Individual and Developmental Differences in Positive Manifold: A Historical and Empirical Investigation of the Differentiation Hypothesis.

Helen Claudia Pagliari

Doctor of Philosophy
University of Edinburgh
1998

Declaration

This thesis represents my own work. I have been personally responsible for writing it and for designing, conducting and analysing the studies reported within it. It has not been submitted in any previous application for a degree.

Acknowledgements

I would like to express my sincere thanks to Dr. Ian Deary and Professor Robert Morris for their patience and continued support during the period of my studies and to Mr. Chris Brand and Dr. Michael Anderson for their past help.

My thanks go, also, to all those friends and colleagues who have stood by me and helped to maintain my confidence and motivation. You have been and will always be appreciated.

CONTENTS

ABSTRACT

This thesis questions whether the positive manifold effect, first observed by Spearman (1925), is equally influential across samples differing in intelligence level and age. In particular, the studies reported here ask whether positive manifold is less evident in older, compared to younger, children and in samples composed of subjects of higher, as opposed to lower, IQ. They also attempt to test Anderson's (1986) theory regarding the role of mental speed in accounting for the strength of Spearman's g.

In the first phase of the research (encompassing studies 1 and 2) approximately five hundred subjects, evenly dispersed between the ages of eight and twelve years, were examined using a battery of ten psychometric tests. The amount of common variance amongst tests was assessed in the five age groups and across IQ ranges. There was no consistent evidence to indicate a reduction in the positive manifold effect with increasing age, after correcting for statistical bias. Similarly, there was no tendency for Spearman's g to be more or less pervasive in groups differing in IQ level.

In the second phase (encompassing studies 3, 4 and 5) subsamples of nine and twelve year olds (c. 60 at each age) were tested with a battery of cognitive tasks thought to access mental speed. These included *Inspection Time* and two response time measures based on *Shepard's Shape Rotation Task* and *Posner's Letter Discrimination Task* (mean RT was the key variable in each). Twelve year olds had significantly shorter response times than nine year olds, for both RT tasks. The older sample also demonstrated significantly shorter inspection times than the younger sample, although this was only the case for the first two of the three estimates obtained. Correlations between IT and IQ and between RT and IQ did not differ significantly in magnitude between the two age groups or between samples of lower, compared to higher, general intelligence. A non -significant tendency was, however, observed for the relationship between inspection time and IQ to be stronger in samples with slower-than-average IT, compared to those with faster-than-average IT. Similarly, correlations between response times for the shape and letter discrimination tasks were non-significantly stronger in the group having slower mean inspection time. No tendency was observed for either measure of response time to correlate more strongly with IQ in the sample having slower IT, however, nor was there any evidence to indicate that RT-IQ correlations were stronger in samples with longer mean response times.

These results do not support either 'differentiation hypothesis', since neither increasing age nor increasing IQ level appear to be associated with a decline in the strength of g. Although Anderson's prediction that RT will decrease with increasing age is confirmed, his theory that basic mental speed (supposedly indicated by IT) does not change during development cannot be said to have been supported. Nevertheless, the possibility that age-related measurement error may disguise true stability in basic cognitive speed cannot be discounted, particularly since the most reliable of the three IT estimates did not show a significant age difference. The non-significant tendency for correlations between IT and IQ, and between the two RT variables, to be stronger in the sample having longer mean inspection time, also allows for the possibility that Anderson's theory of an association between mental speed and g-strength may be correct. Further research, using larger samples and more robust methods for estimating IT, will be necessary if such an effect is to be demonstrated and the theory substantiated. It is concluded, however, that the most likely sources of sub-group differences in positive manifold, where they occur, are sampling and test limitations, although differential learning experiences may also play a role.

PREFACE

Over one hundred years have elapsed since the publication of Sir Francis Galton's pioneering book 'Hereditary Genius' (l869). The intelligence field which Galton's work precipitated has, perhaps more than any other area of psychology, been characterised by continuing academic debate and public interest. Considerable media attention has been directed towards controversial issues such as race and gender differences in IQ, the respective roles of nature and nurture, and the use of psychometric tests for educational selection and streaming. An equally important, 'though less newsworthy, issue has preoccupied and divided theorists within the field since its earliest days. This concerns the concept of intelligence itself; its nature, composition, organisation and operation. This is crystallised in what has been referred to as the *structure of intellect* debate.

Within differential psychology, views on this issue have tended to divide into two broad camps. On one side are those who argue that the full range of intellectual abilities can be understood in terms of a single central trait - general intelligence. On the other side are theorists who believe intelligence to be the product of many independent, specific abilities. Most investigators take a stance somewhere in between these two, conceptualising intelligence in terms of both general ability and specific abilities of varying breadth. Nevertheless, the broad monist/pluralist divide has, historically, tended to separate the so-called 'London' and 'American' schools and continues to be a source of contention.
To a certain extent, these differences can be traced to variations in statistical methods. They are also based, however, on genuine differences in empirical findings. The aim of this thesis is to consider the circumstances which may have given rise to these differences and hence to the theories of the relevant investigators.

The author's interest in this topic was stimulated by a series of articles by Michael Anderson, in which he outlined a model of cognitive architecture which predicts that the relationship between general intelligence (g) and specific abilities changes as a function of the level of basic cognitive ability possessed by the individual (emphasising the role of processing speed). At the time (late 1980s) this appeared to be a novel concept with the potential to greatly improve our understanding of intelligence and account for inconsistencies between monistic and pluralistic theories. Library research for this thesis, however, revealed that the issue had been addressed before, indeed, by the father of modern psychometrics Charles Spearman who, in the 1920s, had proposed a model remarkably similar to Anderson's. Further research uncovered a number of pertinent studies in the international literature, notably work conducted in the United States of America in the 1940s and '50s and in Germany in the 1960s and '70s. The majority of this work had addressed the issue of developmental changes in the structure of intelligence, or the strength of Spearman's g, however much of it implied a strong predictive role for basic intellectual capacity.

A twist of fate set this neglected issue firmly on the agenda of the international intelligence research community. Shortly after the initial literature review for this thesis had been completed and preparations were underway to begin testing subjects, a paper was published in the journal *Intelligence*, by its editor-in-chief, Douglas Detterman and his colleague Mark Daniel (1989). Their paper outlined a reanalysis of national standardisation data for the Weschsler intelligence scales and two smaller comparative studies of young adults differing in IQ level. The results indicated a stronger g-factor in subjects of lower, as compared to higher IQ. The authors hailed these results, and the fledgling theory they put forward to explain it, as a new breakthrough. That these leaders in the field were unaware of previous work on this topic, is testament to the general ignorance amongst intelligence researchers of what had once been, and intermittently has been, an important area.

Subsequent publications, by a number of researchers, appeared in the literature during and after the data collection phase of the project. This was unfortunate, from this author's point of view, particularly since many of these had superior sample sizes. Nevertheless, the results of such studies were mixed, behoving further work. Furthermore, many publications reflected recapitulations of normative data, without addressing the fundamental issues concerning the roots of reported sample differences in positive manifold. The current author has attempted to address these issues, particularly as regards the potential role of statistical biases, in detail. Undoubtedly, a number of subsequent studies have been more rigorously designed and have more statistical power by virtue of their sample sizes. Nevertheless, there remains an important niche in this research market, which this thesis hopes to fill. It is the aim of this thesis to subject the historical, theoretical, methodological and statistical issues associated with the 'differentiation hypothesis' to closer scrutiny than has, heretofore, been done. The studies reported here address both the age-differentiation hypothesis and the ability-differentiation hypothesis. Furthermore, they combine both the structural psychometric approach and the cognitive correlates approach in doing so, the latter addressing Anderson's model of intelligence.

The thesis is structured in the following way:
In the introductory chapters the issue is set in historical context, beginning with a general introduction to the field of intelligence testing and a review of different theoretical conceptualisations of intelligence and approaches to its measurement. The potential impact of sample characteristics (notably age and IQ level) on the structure of intelligence is discussed next and the issue is raised as to whether the historical monist/pluralist divide may be rooted in early findings of different patterns of psychometric ability in different samples.

Chapter 2 describes the existing literature specifically concerning the influence of chronological age on the strength of Spearman's g. Similarly, Chapter 3 is a review of research linking sample IQ level to g-strength. Embedded within the latter chapter is a discussion of converging evidence from the literature on creativity.
Theories which have been put forward to account for differentiation effects are also considered in both of these chapters. Particular emphasis is placed on Anderson's model of minimum cognitive architecture and Spearman's theory of the *law of diminishing returns*.

Chapter 4 details the methodology used in the psychometric studies examining age and IQ differences in g. The results of these investigations are presented and discussed in chapters 5 and 6.

Chapters 7-11 contain the methods and results for the studies employing measures of information processing speed to test Anderson's hypotheses. These are reviewed and discussed in chapter 12.

In the final chapter (13) the key results of the project are summarised and conclusions drawn.

CHAPTER 1
THE STRUCTURE OF INTELLECT DEBATE IN HISTORICAL CONTEXT

The emergence of the intelligence field

The first serious attempt to subject intelligence to scientific study was made by British theorist Francis Galton in the late 1860s. In an effort to determine to what extent ability is learned or inherited, Galton identified key persons of "eminence" in past and present history and assessed the achievements of their biological relatives. His findings of "hereditary genius" (l869) were supported by the biologist Herbert Spencer (l870) and the writings of these two scholars can be credited with introducing the term 'intelligence' to the popular vocabulary.

Early attempts to measure intelligence in the laboratory

Galton's biological perspective led him to speculate that intelligence is primarily the product of a global, general capacity, related to the number, complexity of connections and organisation of the nerve cells in the cerebral cortex. He reasoned that these elements would be tapped in measures of sensory discrimination and, in the late 1870s, he set up a laboratory with the aim of measuring individual differences in such skills and relating these to measures of complex abilities normally associated with academic achievement. Although Galton reported a positive relationship between these two types of variable, his tests were too crude, and his observations too subjective, to yield any firm conclusions. Nonetheless, his research was fundamental in stimulating debate about the underlying nature of this complex trait.

On the other side of the Atlantic, a similar programme of research was underway at Columbia University, under the direction of James McKeen Cattell (1890). Cattell measured the performance of college students on a series of "mental tests" consisting of simple measures of motor skills and reasoning ability. Clark Wissler, working in the same laboratory, used correlational methods to assess the degree of relationship between these variables and, in a landmark paper of 1901, he reported that there was no consistent correlation between such tests or between the simple tests and measures of academic performance. Wissler's paper was widely disseminated and his conclusions were largely responsible for the decline of interest in this form of intelligence testing in the United States. His work also set the trend towards theories in which intelligence is conceptualised in terms of specific, independent processes.

Initially unaware of Wissler's conclusions, British psychologist Charles Spearman had embarked on a series of investigations building on Galton's earlier work. It is largely Spearman's findings, methodology and theories which have provided the basis for modern psychometric theory.

Spearman (1904) tested his subjects (mostly schoolchildren) using a number of measures of sensory discrimination (visual, auditory, tactile). He also collected their marks on a variety of academic examinations and teachers' subjective ratings of their intelligence. He quantified the relationships between these variables using correlation coefficients.

Spearman's discovery and the development of the theory of g.

Spearman's key finding was that performance estimates for every one of the variables were positively intercorrelated. This led him to theorise that all tests measure a single central trait, to a greater or lesser extent. This general ability trait he termed "g", and the phenomenon of ability covariation he called *"the law of positive manifold"*.

Like Galton, Spearman believed that this global trait is a product of biological features of the brain which influence its efficiency. He also accepted the idea that the brain is composed of many parts which are specialised for different functions, but he believed the general trait to be of primary importance in explaining individual differences in all of these.

In his later mechanical model of the mind, Spearman (1925, 1927) likened intelligence to a factory composed of a central furnace and specific, task-dedicated, engines. Factories (individuals) vary with respect to the energy output; or power; of their central furnaces. The specialised engines also differ in quality between factories, however the primary constraint on their performance is the energy which is available to them. This is determined by i) the overall power of the furnace and ii) the competing demands of the other engines at any one time. Because it influences all other parts of the system the furnace gives rise to g. More significantly, factories with more powerful furnaces will be better able to deal with tasks which involve many engines simultaneously. In other words, because humans vary with respect to their central resources, they differ in the range of cognitive equipment available to them for solving mental tasks. The capacity to hold several ideas in mind and make rational connections between them or, to use Spearman's phrase, *the* *eduction of relations and correlates*, is the essence of intelligence. Spearman's model thus shows how individual differences in one central variable will yield performance differences in both simple and complex tasks.

According to Spearman's *'two-factor theory'*, tests vary in their capacity to measure g (i.e. their correlation with all other tests in the universe of tests) and in their specific features (which include measurement error). The correlation between any two tests is a function of the ratio of g to s in the respective tests. This *'law of tetrad differences'* can be written as follows:

$$r_{12} \times r_{34} = r_{13} \times r_{24} \qquad \text{(Spearman \& Holzinger, 1924)}$$

Since all tests tap g, but individual tests vary in their capacity to do so (their g-loading), an individual's g-level will be most reliably indicated by the aggregate score on a battery of diverse measures. Spearman reasoned that if enough tests are used, g will be revealed, irrespective of the specific composition of the test battery. This he referred to as *the indifference of the indicator* and the use of an aggregated score he called the *hotchpotch principle.*

An early challenge to the theory of *g*.

One of Spearman's main contemporary challengers was Thomson (1916) who maintained that Spearman's data could be better explained by an 'anarchic' model of the mind than by the 'monarchic' theory of g. Thomson theorised that the mind is composed of a set of independent units or 'bonds'. Any task will involve a subset of these bonds, hence the capacity of a test to detect generalisable ability differences is a function of the breadth of bonds it samples. In other words, the statistical loading of a test on the g factor (essentially the degree to which the test correlates with all other tests) is reflective of the breadth of its bond-sampling, rather than its ability to tap a central energy/efficiency variable. Thomson also believed that general intellectual differences between individuals result from differences in the number of bonds they possess. Thus, a person having a greater range of bonds at their disposal will be able to solve tasks of greater complexity.

As has been pointed out by Brody (1992), the models of Spearman and Thomson are not incompatible, their main difference being metaphorical. Both account for general intelligence. Nevertheless, the debate over whether g is a product of one entity or of the coaction of many, preoccupied the two theorists for over twenty years.

The emergence of psychometric testing and the IQ.

Spearman's early work was instrumental in developing a theoretical basis for applied intelligence testing. His initial laboratory-based methods were impractical for the purposes of educational selection and guidance, however. (They also imposed constraints on the number of subjects who could be tested, restricting the evidence on which to base further theoretical developments.) Mindful of the need for valid, reliable and easily administered assessment devices, the French ministry of Education in 1905 commissioned psychologist Alfred Binet and his associate Theodore Simon to develop a set of tests for diagnosing mental retardation. The measure they developed can be thought of as the first intelligence test, in the modern sense.

Binet accepted the notion of general intelligence, defining it variously as *"judgment, good sense, practical sense, initiative, the faculty of adapting oneself to situations"* (Binet & Simon, 1905). Like Thomson, however, he believed it to result from the combined influence of several independent mental functions, or 'faculties', rather than one central trait. He argued that efforts to measure central traits through simple laboratory measures were of limited value, since this approach did not address the multi-faceted nature of intelligence. Nevertheless, he maintained that since most intellectual tasks involve some combination of specific functions, complex tasks involving several faculties are likely to be most useful for the purposes of educational selection.

Binet & Simon's scale was designed around the idea that children's problem solving skills improve with age. By identifying intellectual problems which the majority of children in one particular cohort will be able to solve, each 'mental age' (MA) can be operationally defined. A child whose capacity to solve these problems deviates from that of the majority can be classified as having a mental age below or above their chronological age. The test itself was composed of 30 items arranged in order of increasing difficulty. The child's MA was estimated according to the last item to be solved, with all items having been scaled for age-equivalence.

Having shown that such a scale could be useful for the diagnosis of mental retardation, Binet and Simon went on to develop a more widely applicable test for normal children (Binet,1908,1911). This was translated into English by Lewis Terman (1916) and the product became known as the Stanford-Binet test. The classification of testees by simple mental age was later abandoned in favour of the "mental quotient" originally suggested by German psychologist Wilhelm Stern (1907), in which the individual's performance is expressed in terms of a ratio of MA to CA . Terman renamed this the 'intelligence quotient' or IQ (IQ = MA/CA x 100). In modern tests this method of calculating IQ has been superseded by Wechsler's (1958) 'deviation IQ' which calculates the individual's performance level relative to a standardisation sample whose scores are normally distributed. The population mean and standard deviation are nominally set at 100 and 15 respectively and an individual's distance from the mean can thus be expressed in IQ points.

Early variations in the psychometric structure of intelligence.

Following Binet's success, interest in mental testing burgeoned and a large range of psychometric measures were constructed, ranging from those covering a relatively broad range of abilities to those measuring more specific skills. The data which emerged from this movement provided greater opportunities to examine the structure of intelligence. Not surprisingly, Charles Spearman made extensive use of Binet-type tests, and the theories outlined earlier are derived largely from his work using similar measures.

Thurstone's Approach

The first major challenge to Spearman's position, from the psychometric field, came from the American psychologist Louis L. Thurstone. Like Thomson and Binet, Thurstone (eg.1938) believed that Spearman's theory of g was too parsimonious to capture the full diversity of intelligence. He theorised that intelligence is a function of several independent abilities and he argued that an adequate decription of a person's intelligence is only possible by assessing their performance on each of these individually. Thurstone's main argument was empirical. Like others before him, he realised that Spearman's data did not perfectly support the law of tetrad differences: There were some tasks which correlated more with each other than would be expected from their g-loadings and tests with equally high g-loadings which were not themselves highly correlated. Spearman had argued that the former only occurred where the tests were very homogeneous (producing what Cattell later called 'bloated specifics') and the latter were a result of measurement error. Thurstone, however, interpreted such results as evidence that intelligence is multidimensional and that these dimensions are independent. He argued that direct methods of factor analysis (such as used by Spearman) are limited in their capacity to reveal specific factors, since they account for most of the variance in a matrix in terms of g. To rectify this, he designed a new method to more clearly identify the psychological dimensions which underlie observed patterns of correlations in a matrix. In *multiple factor analysis* , the matrix is first examined for the presence of clusters of variables which correlate highly with one other but not with other variables. This gives a rough guide to the type of ability trait underlying the cluster. The reference axis for each of the hypothesised factors is then rotated in such a way that the loadings of the majority of variables on that factor are at or near zero, but the positive correlation of one test with the factor is very high. The factor is thus defined by the test which loads on it to the greatest extent. The state of maximum factor delineation Thurstone referred to as *simple structure*. Thurstone (1938) used this method to analyse the scores of 240 university students on 56 tests. Initial analysis suggested the presence of thirteen different factors. Nine of these were labelled psychological abilities: spatial, perceptual, numerical, verbal, memory, word fluency, inductive, arithmetical reasoning and deduction.

Although Thurstone had started out by insisting that the factors should be uncorrelated, or orthogonal, satisfactory simple structure was hard to achieve. Positive correlations between factors implied the presence of Spearman's g. Thurstone initially resisted this possibility, arguing that the results had to do with task complexity, but he was eventually forced to concede that the abilities which he had identified shared common variance. He also realised that the very low inter-test correlations which he had initially observed resulted, to a large degree, from sample range-restriction. Simple structure was even more elusive when a more heterogeneous group of subjects was used.

By 1941 Thurstone had relinquished his criterion that the factors must be uncorrelated and had modified his technique to allow for correlated (oblique) factors. The correlations between these factors could themselves be subjected to second-order factor analysis. He eventually concluded (Thurstone & Thurstone, 1941) that intelligence comprises roughly seven "primary mental abilities" (PMAs), with respect to which all test variance can be explained, along with a second-order general factor (probably equivalent to Spearman's g). Eight subtests are included in the 1963 edition of the PMA, three of which are sufficiently close to others to merit a 5 factor classification system. The five scales cover the domains of verbal comprehension, numerical ability, spatial ability, abstract reasoning and perceptual speed.

Thurstone's research was of great value to the mental testing movement, in that it in directed attention beyond the g factor. Later investigators went further in their attempts to uncover abilities of increasing specificity. Their aims were both to improve knowledge about intelligence by understanding its building blocks and to produce tests with the maximum capacity to predict specific occupational or educational outcomes. Investigators have varied widely in the number of ability factors which they claim to have isolated, and in the degree to which these are held to be independent.

Guilford's Model

The theorist who claims to have demonstrated the greatest number of specific ability factors is J. P. Guilford. Guilford's "structure of intellect" (SOI) model, posits the existence of one hundred and fifty independent abilities (Guilford, 1967; Guilford and Hoepfner, 1971; Guilford, 1982). This number is arrived at by describing tests and factors in terms of three facets: *operation*, *content* and *product*. There are five kinds of operation: cognition, memory, divergent production, convergent production and evaluation; five kinds of contents: visual, auditory, symbolic, semantic and behavioural; and six kinds of products: units, classes, relations, systems, transformations and implications. Since the sub-categories are independently defined, they are multiplicative so there are 5 x 5 x 6 = 150 different mental abilities (a "cubic" factor arrangement).
As of 1982, Guilford claims to have isolated 105 of the 150 possible factors. As was the case for Thurstone, however, his original prediction that these would be uncorrelated has not been supported by the data, behoving him to revise his model to one which is not incompatible with the theory of g (although it should be pointed out that Guilford himself is resistant to this description).

At this point, it is important to re-emphasise that differences between monist and pluralist theories of intelligence arise at both the *empirical* and the *conceptual* level. Thorndike (1994), for example, distinguishes between g as a universally positive correlation between cognitive tasks and g as a psychological construct. Similar empirical findings may, therefore, be interpreted differently by researchers having variant theoretical perspectives. Today's theorists remain divided on the issue of whether the root of intelligence is one central, pervasive trait, or whether it arises as a result of several independent (but often co-acting) abilities. These divisions are based on both conceptual and empirical grounds.

Later views of intelligence

Douglas Detterman (e.g. 1987) sees intelligence as a complex *system* composed of several independent abilities. These vary in the degree to which they are important for the functioning of the system as a whole, or the number of mental tasks in which they are implicated. This affects their statistical loading on the general factor. Like Spearman and others, Detterman believes that complex tasks involve the coaction of many specific abilities. Unlike Spearman, however, he sees g as a by-product of this coaction, rather than as a cause.

Other theorists prefer to look beyond the psychometric structure of abilities in attempting to explain intelligence. Robert Sternberg's theories, for example, have concentrated on the individual, rather than on individual differences. According to his *triarchic theory* of intelligence (Sternberg,1985) it is necessary to explain intelligence at three levels: Explanations at the *contextual* level take account of the environment in which the individual is operating, explanations at the *componential* level involve the specific cognitive processes underlying performance and a third type of explanation (*two-facet* theory) involves the associations between the individual's internal and external worlds.

Howard Gardner (1983) is another theorist who has departed from the structural psychometric approach in explaining intelligence. Gardner has taken the pluralist perspective to the extreme, arguing for the existence of six independent *intelligences* : musical, linguistic, logical mathematical, spatial, bodily-kinesthetic and personal (self understanding and empathy). He claims that the existence of these independent intelligences is indicated by the following evidence:

1. Findings of isolated ability deficits following brain damage.
2. Idiot savants - Individuals who are exceptionally gifted in one or more abilities but weak or mediocre in the majority.
3. A core set of mental operations underlying each type of intelligence (e.g. pitch discrimination underlies his hypothesised musical intelligence).
4. A clear developmental history.
5. Evolutionary plausibility.
6. Support from experimental psychological tasks demonstrating the independence of specific mental operations.
7. Psychometric findings indicating trait independence.
8. Symbolic encoding systems (e.g. vocabulary, syntax and grammar characterise linguistic intelligence).

Despite its intuitive appeal and early popularity, Gardner's theory is without a firm evidence-base and his choice of 'intelligences' has been criticised as arbitrary (e.g. Brody,1992).

A focus on the *processes* underlying intelligence, rather than on the psychometric structure of intelligence, has characterised much recent research in the field. A number of investigators have used mental chronometric methods, from experimental psychology, to try to elucidate the cognitive sub-processes characterising performance on particular types of test, such as those measuring spatial or verbal abilities (e.g. Hunt, 1978; Shepard & Metzlar, 1971 and S. Sternberg, 1970; R. Sternberg, 1977, 1985).

Robert Sternberg (1977), for example, attempted to delineate the cognitive sub-processes (or components) involved in solving analogies, such as *inferring* the relation between the first two terms, *mapping* the higher order relation that connects the first half of the analogy to the second, and *applying* the relation inferred in the first half of the analogy to the second half of the problem. Although Sternberg has tried hard to steer attention away from g, he has repeatedly found there to be a positive correlation between measures of performance on diverse cognitive tasks. This has led him to postulate 'metacomponents' of intelligence which are not incompatible with Spearman's thesis.

It has become common practice to integrate the experimental and psychometric approaches by examining the psychophysical correlates of IQ. Investigators following this approach have tended to take a strongly reductionist stance, their primary aim being to identify the biological (possibly inherited) basis of g (Eysenck, 1988). Like the early work of Wissler and McKeen Cattell, much of this research has been concerned with measures of general mental speed or efficiency, such as *reaction time* , *inspection time* and measures of electroencephalic activity (e.g. Jensen,1987; Nettelbeck, 1987; Hendrickson & Hendrickson, 1980). This is typical of the monist approach in that the main interest lies in identifying a single explanatory variable for intelligence. [This research will be described in more detail in a later section.]

Hierarchical Models

Returning to psychometric studies of the structure of intelligence: It will have become clear, from the discussion thus far, that from an early stage neither a strictly monist or a strictly pluralist perspective has been sufficient to explain observed patterns of inter-test correlations. Hierarchical models of intelligence, which encompass factors with varying degrees of generality, offer one means of reconciling the two camps. The value of such models was recognised as early as 1941, when R.B. Cattell suggested that the findings of Spearman and Thurstone could both be accommodated by postulating a hierarchical structure of ability. As mentioned earlier, differences in reported factor structure are often a result of different methods of analysis, with indirect methods tending to produce more clear-cut primary factors. Indeed, when Eysenck (1938) reanalysed Thurstone's data using direct factorial methods, he found that between thirty and fifty percent of the variance could be explained by g. He also found that after g had been extracted, eight factors remained which corresponded roughly to those which Thurstone had identified.

Three main hierarchical models have been proposed which, although they are derived using different techniques of analysis, are broadly compatible.

The first of these may be termed the Burt/Vernon model. Burt (1949), Vernon (1950, 1965), and others (e.g. Harman 1967), developed hierarchical group factor techniques which, from a matrix of correlations, extract first the g factor, and then group factors of successively smaller breadth. The Vernon model has as its pinnacle the g factor. At the next level are two broad group factors, labelled *verbal-educational* (v:ed) and *spatial-mechanical* ability (k:m). The v:ed factor subdivides into narrower scholastic factors such as vocabulary and number abilities and the k:m factor into perceptual, spatial and mechanical factors. Each of these narrow factors can be further subdivided by more detailed testing.

Cattell and Horn (e.g. Cattell and Horn l966) have used a different approach in developing their hierarchical model. It is based on oblique multiple factor analysis of several orders. In contrast to the Burt/Vernon method, the first step is to produce a large number of first order factors from the correlation matrix. The correlations between these factors are then subjected to further factor analysis to yield a number (approximately five) of second order or general factors. Of these, the two most important are usually *fluid intelligence* (g f) and *crystallised intelligence* (g c).
Fluid intelligence is characterised as biological or innate aptitude. Tests which measure g f most effectively involve novelty, or the manipulation of complex and abstract symbolic information (e.g. Raven's Progressive Matrices). This is similar to Spearman's concept of g as the product of fundamental qualities of the individual's central nervous system which yield differences in the capacity to educe relations and correlates. It is primarily g f which is thought to be accessed by measures of so-called 'biological' intelligence such as reaction time (e.g. Jensen 1982), inspection time (eg. Brand & Deary 1982) and the evoked potential (e.g. Hendrickson & Hendrickson,1982).
Crystallised intelligence in contrast, is the result of learning and experience, hence tests which measure g c most effectively involve learned information. The capacity to learn (g f) and the product of that

capacity (g c) are, of course, related: It is difficult to measure g f without relying on some form of learned information and g c is a reflection not only of what is learned, but what it has been possible to learn given the constraints imposed by biology and inheritance. Nonetheless, it is possible to devise tasks which make demands predominantly on either g f or g c. [Fluid and crystallized intelligence correspond to what Hebb (1946) later called *intelligence A* and *intelligence B*.]

Gustafsson (I984, 1988) has provided a hierarchical model which encompasses both previous models, and is widely acknowledged to be the most useful of the three. The model was derived using confirmatory factor analysis, by which the inter-variable relationships which are predicted by different models, can be tested for goodness-of-fit to the relationships which are actually found in the data.
General intelligence is at the apex of the hierarchical structure identified by Gustafsson. This general factor is closely identified with Cattell's fluid intelligence (g f). At the next level of the hierarchy are two broad group factors: *verbal-educational* ability (similar to g c) and *visuospatial* ability. These are equivalent to those labelled v:ed and k:m by Vernon and correspond roughly to the functions of the left and right cerebral hemispheres, respectively. Each broad group factor is then divided into narrow group factors, roughly corresponding to Thurstone's PMAs and a number of other similar level factors. These can be further subdivided until specific factor level is reached. It is plausible to assume that these specific factors may correspond to Guilford's I50 abilities. (Guilford's revised model does, in fact, conceptualise the structure of intelligence in hierarchical terms, although it does not include a superordinate g factor.) Gustafsson's model has been shown to fit the diverse data most completely and is (according to Lynn, 1987) the structure of intellect

model which commands general acceptance among knowledgeable researchers in the field.

Might the monist/pluralist divide have an empirical source in sample characteristics?

While it may be possible to subsume a number of diverse psychometric models within a hierarchical framework, the variation between monist and pluralist accounts of intelligence cannot be fully resolved in this way. For one thing, the same hierarchical data structures may be interpreted in different ways by observers having different theoretical perspectives (for example, general factors may be seen either as the cause or consequence of specific ability variance.) Likewise, hierarchical analyses cannot, themselves, overcome the subjectivity inherent in factor definition. More importantly; from an empirical point of view; the extent to which Spearman's g has been found to account for the total variance in matrices of inter-test correlations (and, consequently, the number of factors extracted) varies between studies/data sets. This may have implications with regard to the foundations of the monist and pluralist schools.
As implied earlier, the process of developing models of intelligence is both top-down and bottom-up. This is to say, research findings may emerge as a result of the theoretical perspective held at the outset and the associated methods employed to test the theory. Similarly, the theory may itself emerge from the data. It is therefore important to consider whether the differences between monist and pluralist accounts of intelligence may have an empirical source and, if so, why.

This is the key issue addressed in this thesis. Of primary concern is the question of whether the strength of Spearman's g differs across samples. Such a possibility is suggested by a comparison of the *subjects* used in the early American and British studies already described. The data of Wissler, Thurstone and many other early American researchers were largely derived from the testing of college students. In contrast, the samples tested in early British studies, notably Spearman's, were more commonly composed of school children. This is intriguing insofar as the findings of these researchers propagated models emphasising multiple abilities and general intelligence, respectively. As will be discussed in the following chapters, various investigators have, over the years, speculated that in older, brighter and more selected subjects, the g factor decreases in importance and specific abilities assume a greater explanatory role in accounting for individual differences in intelligence.

The issue of sample selection is a statistical one, which Spearman recognised as early as 1904. In short, inter-test variation decreases with the ability range in a sample. This has the effect of lowering correlations, yielding a weaker g-factor when the matrices are factor analysed. College students are highly selected for ability and it has already been pointed out that Thurstone partly attributed his initial findings to this fact. School children are less likely to be as homogeneous in terms of ability. Some weight must therefore be given to the possibility that early British/American differences in factor structure were a by-product of range restriction effects. This argument has been put forward by writers such as Vernon (1965). Indeed, Guilford (1969) has even attempted to explain away the existence of g as a statistical artifact arising from sample heterogeneity.

Of far greater implication for the theory of positive manifold on which most modern theories rest - and of central importance to this thesis - is the possibility of greater ability co-variance (i.e. a stronger g factor) in *younger* and in *less intelligent* subjects. Observations of such a pattern have been reported frequently over the years, and have led to a number of hypotheses concerning the structure and development of intelligence. Amongst the labels used to describe these, the term *differentiation hypothesis* is fairly representative.

Despite their potential implications for both the theory and application of intelligence testing, hypotheses of differentiation have been given relatively little research attention in recent years. Although the results of many studies support such hypotheses, few have been conducted to specifically test them. With respect to the influence of *intelligence level* on the strength of Spearman's g, the literature is very limited. It is not surprising, therefore, that even the chief editor of the flagship journal *Intelligence* was unaware of previous work when he published his renaissance paper, shortly after the outset of this thesis (Detterman & Daniel, 1989).
Studies which have addressed the issue of *age* differences in the strength of Spearman's g are more numerous. Nonetheless, there is little knowledge of this work among modern investigators - a situation which can probably be attributed to the fact that most of the relevant papers were published prior to 1960.
In view of the importance of this issue to the structure of intellect debate, and its implications for applied intelligence testers, it is an area which greatly requires further investigation.

The aim of the following two chapters is to review the literature in support (or otherwise) of the "Differentiation Hypotheses"; hypothes<u>es</u> because one is related to specialisation with age and one to specialisation with increasing intelligence. As will become apparent, however, it may be that the two can be accounted for within one theoretical framework.

CHAPTER 2

DEVELOPMENT AND THE DIFFERENTIATION OF ABILITIES

Early Theories

Spencer's 'Developmental Hypothesis'

It was Herbert Spencer who, in 1870, first presented a theory pertinent to the differentiation hypothesis. His 'developmental hypothesis' associated the intellectual development of the child with the evolution of the species. Spencer theorised that like lower, as compared to higher, organisms, younger children have more simple, undifferentiated, intellectual capabilities than do older children. Maturation is accompanied by neurological specialisation which, in conjunction with specialisation due to learning, leads to a diversification of abilities. The fully developed human brain, therefore, exhibits a hierarchical organisation, its intellectual functions ranging from simple perceptual skills to complex thought.

Whilst Spencer's evolutionary thesis may be questioned, his theory, that intellectual development is characterised by a shift from simple, undifferentiated skills to complex cognitions, is compatible with those of many later theorists including Binet, Piaget and others of the French and Swiss schools.

Binet's theory of hierarchical development

As described in the last chapter, Binet (1905) accepted the notion of general intelligence. He believed, however, that *"at different stages of mental development, this manifests itself in various ways. There is, in fact ...a hierarchy among the diverse manifestations of intelligence."* According to Burt (1954) Binet's view was essentially that of Spencer, namely that *"the more specific abilities are differentiated progressively as the child matures: they are thus distinguishable, though not separate, from general intelligence".*

Piaget's theory of progressive differentiation

Piaget (1950, 1953), also adhered to the Spencerian doctrine of the hierarchical development of intelligence. In his chapter on "The Hierarchy of Operations and their Progressive Differentiation", he observes that *"each of the transitions from one level to the next is characterized both by new co-ordination and by a differentiation of the systems of the preceding level: at the same time they illuminate the undifferentiated nature of the initial mechanisms".*

Burt's Differentiation Hypothesis

The first theorist to explicitly hypothesise that the factorial structure of intelligence will differentiate with increasing age was Cyril Burt (1918). Like Spencer and Binet, Burt believed that the brain becomes more specialised in function during child development, both as a result of biological maturation and the aquisition of new knowledge. Consequently, he theorised, the strength of the general factor will decrease and specific factors will assume a greater role in explaining intelligence. This prediction was confirmed in a set of experiments described in Burt's paper of 1954. These will be described in the next section.

Spearman's Law of Diminishing Returns

Charles Spearman recognised the implications of Burt's findings for his own theory of positive manifold and, in 1925, proposed a theory to explain it. As described earlier, Spearman conceived of intelligence as a system composed of a central energy source responsible for g, and a set of semi-independent engines associated with specific abilities. His theory of differentiation will be described more fully in a later section. In short, it suggests that ability divergence is a result of both maturation and knowledge aquisition. Maturation is associated with a gradual increase in general mental 'energy'. In early childhood the performance of all the engines is constrained by the amount of energy available. As age increases, so does "the energy". After a certain point, further additions of energy do not lead to corresponding increases in performance. This Spearman referred to as the Law of Diminishing Returns. With ample energy resources the engines are capable of functioning to their full potential, hence their differences in potential become clear. This is the main reason, theorises Spearman, why ontological development is associated with a gradual diminution of positive manifold. The differentiation effect is ehanced because as skills become well learned, they become less reliant on the 'energy' (central resources, or g) and are more accurately predicted by the specific 'engines' (specific abilities).

Garrett's Developmental Theory

American researcher Henry Garrett put forward his own differentiation hypothesis in 1938, although he appears to have been unaware of Burt's work, or Spearman's explanation, at the time of doing so. According to Garrett "*...intelligence changes in its organization as age increases, from a fairly unified and general ability, to a loosely organised group of abilities or factors"*. (l938, p.373). Support for this hypothesis was derived from a series of studies conducted at Columbia University, the results of which are reviewed in his landmark paper of 1946 entitled 'A Developmental Theory of Intelligence'. While Garrett attributed ability differentiation, in part, to biological maturation, he also believed experience to play a major role, with greater competence on specific tasks, and the tendency for similar skills to be acquired together, leading to better factor definition. He also believed test selection to have a role, since older children are able to complete a wider (more differentiated) variety of tests.

The following section will describe the relevant studies of Burt, Garrett and others concerned with the age differentiation hypothesis. This will be followed by a discussion of other theories which have been put forward to explain the findings.

Testing the age differentiation hypothesis: A summary of studies.

The first study specifically designed to test the age differentiation hypothesis was conducted by Cyril Burt in 1918. Burt gave a series of tests to *"socially and intellectually representative samples of children at each successive year of life"* and factor analysed the resulting inter-correlations. He discovered at every age both a general factor and several group factors but he also noticed a marked tendency for the group factors to become increasingly predominant with increasing age. He concluded :

> "... the relative importance of the more general capacity (g), is far greater in earlier years as contrasted with later. With younger children, one can often demonstrate little but the influence of the general factor; with older children, and particularly with college students, little but specific abilities or specialised interests. During early childhood, specialization the exception rather than the rule." (Burt, 1954, p. 80)

Many of the tests which Burt used in his early studies were measures of scholastic achievement. He acknowledged, therefore, that the changes in factor structure which he had observed, may have reflected merely the influences of increased experience, learning and diversity of interests. A study utilising standardized aptitude tests, however, confirmed the findings (Burt, 1954). Three hundred boys were examined with tests of 'special aptitudes' at the age of 9 - 10 years and then again, four years later. On both occasions, the same tests and the same subjects were used. At both ages, statistical analysis revealed six distinct factors: Firstly, general cognitive ability or g; secondly, two broad group factors corresponding approximately to fluid and crystallized intelligence; and thirdly, three minor group factors - a memory, a perceptual and a motor factor. As Burt points out:

> "What is important ... is not so much the nature of the factors as their relative

> predominance.... during the period in question, the influence of the general factor appreciably declines (from 53% to 48%), while the influence of the group factors increases to nearly twice its original size (from 28% to 47%)." (1954, p.82)

Burt interpreted his findings as supporting the developmental models of Spencer and Binet. Based on these theories and his findings, he concluded:

> "...in order to measure the intelligence of a particular child, it is unnecessary, and indeed impossible, to test every type of intellectual process: with the younger or more defective, we can measure manifestations on the lower levels only; with the older and the brighter, we shall need to measure higher mental processes." (1954, p. 77)

Spearman (1925) cites findings by Burt of the following correlations between a test of general reasoning and teachers' estimates of pupils' intelligence at four different ages:

Age	10-11	11-12	12-13	13-14
Correlation	.78	.81	.64	.59

He also compares the correlations reported by Otis between specific aptitude tests and g in a sample of school children, with those reported by Carothers for a sample of university students tested with very similar measures:

Test	Correlations with g	
	Otis, grades IV-VII	Carothers, students
Analogies	.84	.71
Completion	.88	.53
Directions	.86	.45
Digits, memory	.41	.22

Both studies indicate that the general factor is weaker in older, as compared to younger, subjects. The latter comparison should be interpreted cautiously, however, since the older subjects had undergone a greater degree of selection.

Garrett (1946) reports the results of a programme of research initiated in 1938. His evidence is derived from a comparison of results from previous studies whose subjects varied in age (Garrett, 1938; Thurstone 1938; Thurstone and Thurstone, 1941), as well as from four studies specifically designed to test the age differentiation hypothesis. In the latter investigations, primary school children at two or three age levels were examined with the same tests. Three of these studies were cross-sectional, (Clark, 1944; Garrett, Bryan and Perl, 1935; Reichard, 1944) and one was longitudinal (Asch, 1930).

Garrett, Bryan and Perl (1935) tested groups of school boys and girls at three age levels using ten measures of memory, verbal and number abilities. In all, 646 subjects were tested, 225 at age 9, 196 at age 12 and 225 at age 15. Considerable effort was made to obtain comparable samples. With one exception, the inter-correlations among memory, verbal and number tests showed a regular tendency to decrease with age from 9 to 12, and from 12 to 15. The average inter-correlation at ages 9, 12 and 15, was, for boys, .30, .21 and .18; for girls, .27, .30, and .10. A multiple factor analysis of the correlations at each age level (boys and girls kept separate) substantiated the correlational evidence. The proportion of variance accounted for by the first unrotated factor (roughly equivalent to g), at ages 9, 12 and 15, was, for boys .31, .32 and .12; and for girls .31, .24, and .19.

Clark (1944) administered Thurstone's Chicago tests of Primary Mental Abilities to 320 boys, roughly 100 each at ages 11, 13 and 15. (Subjects were matched for socioeconomic level). Inter-correlations between scores in each of the six primary factors, V, N, S, W, M and R, showed a regular tendency to drop with age. Average factor correlations (excluding M, which was very unreliable) were as follows: At ages 11, 13 and 15; N and the battery, .62, .55 and .49; S and the battery .48, .47 and .35; V and the battery .62, .55 and .49. These results confirmed the existence of the second order general factor found by the Thurstone's (g) and show, moreover, that it gradually weakens with age. Each group had the same distribution of IQ, ruling out the possibility that the differences in g-strength were due to selection effects.

Clark's findings were partially supported by Reichard (1944) who found overall inter-correlations between verbal, numerical and spatial test scores to decrease sharply from age 12 to age 15, (from .43 to .38 for boys, and .51 to .37 for girls). From age 9 to age 12, however, inter-correlations rose appreciably for both sexes, but this discrepancy was attributed to an unrepresentative 9 year old sample.

Asch (1936), re-tested a group of 189 twelve-year olds who had previously been tested by Schiller (1934) at the age of 9 years, using a battery of twelve tests of numerical, spatial and verbal abilities. Asch found that by age 12, average inter-test correlations had dropped appreciably from .56 to .41 for boys, and from .59 to .51 for girls.

Further support for the age differentiation hypothesis comes from a comparison of a number of studies by L.L. and T.G. Thurstone. As noted in an earlier section, L.L. Thurstone's (1938) paper *'The Primary Mental Abilities'* was based on the results of a study in which 56 tests were given to 240 college students ranging in age from 16 to 25. The inter-correlations among the primary factors were found to be negligible, the median correlation being .03 and the largest .24. These yielded almost the same factorial matrix whether the transformations were oblique or orthogonal. Thurstone concluded, therefore, that the primary factors extracted were essentially independent. A later study using 1154 eighth grade children, however, yielded inconsistent results (Thurstone and

Thurstone l94l). Centriod factor analysis of sixty tests involving words, numbers, spatial problems, diagrams, dot patterns, pictures and mazes, yielded ten factors, later reduced to six primaries. Test and factor inter-correlations, however, were substantially higher than those observed in the previous study of adults, and from the correlation matrix emerged a general factor. This was described as a *"second-order general factor ... probably equivalent to Spearman's g"*. Although such differences could plausibly have resulted from restriction of range in the adult sample, the findings of such different factor structures in children and adults led the Thurstones to investigate further. In l954, they published the results of a study which supported a hypothesis of progressive differentiation of abilities with age. Thurstone and Thurstone (l955) correlated four of their PMA test variables in school grades one and ten and also in a group with grades seven to eleven combined (samples were matched for ability range). Average inter-correlations were .64, .22,and .27 respectively, indicating a marked decrease from the first grade of primary school to high school.

Balinsky (l94l) factor analysed scores on the Wechsler Bellevue test battery, obtained by standardisation samples aged 9, l2, l5 and 25-9. The percentages of variance attributable to g were, respectively, 38, 36, 24, 20, supporting the hypothesis of progressive ability differentiation from childhood to adulthood.

German psychologist G.A. Lienert became interested in the differentiation hypothesis during the early 1960s. Lienert and Faber (1963) found little evidence to support it, but Lienert & Crott (1964) reported a decrease in the proportion of variance accounted for by the first centroid factor, from 48% at age 10-12 to 45% at age 18-20.

Oerter, Mandl & Zimmermann (1974) tested 1511 second grade school children using a German intelligence test battery (the BT 2-3) and the Primary Mental Abilities tests. Subjects were re-tested the following year. The proportion of variance explained by the first unrotated factor (an indicator of g-strength) dropped from 42.40 to 41.50 for the PMA tests and from 47.50 to 43.90 for the German test battery. The authors speculate that this change is the result of learned strategies for problem solving.

Atkin, Bray, Davidson, Herzberger, Humphreys & Selzer (1977) compared samples of children tested at American school grades 5, 7, 9 and 11 using sixteen measures. These ranged from tests of scholastic achievement (e.g. mathematics, social science), information (e.g. home arts, music) and aptitude (e.g. verbal, quantitative). Samples were drawn from the growth data collected by the educational testing service (Hilton et al., 1971) and the study design was retrospectively longitudinal. From a total sample of around 10,000 subjects, a full set of 16 scores was only available for a proportion. The study sample comprised 668 white males, 762 white males, 172 black males and 215 black females. To assess the presence or absence of differentiation, Atkin et al. calculated latent roots from the matrices of inter-score correlations in the four groups at the four grade levels. The strength of Spearman's g in each sample is indicated by the size of the first latent root. These figures are given below:

Size of the first latent root derived from 16 measures at 4 ages (Atkin et al.,1977)

	Grade 5	Grade 7	Grade 9	Grade 11
White males:	10.17	9.11	8.88	8.92
White females:	9.41	8.86	8.53	8.84
Black males:	8.43	6.99	7.50	7.73
Black females:	8.77	8.10	8.36	8.23

These results indicate a decrease in the strength of Spearman's g with increasing age. Atkin et al. also report a small increase in the number of factors emerging from the matrices, again supporting the hypothesis of progressive ability differentiation.

Despite the considerable evidence in favour of the age differentiation hypothesis (e.g. Chuprikova,1990) it is not, as Butcher (1968) would have us believe, *"generally accepted"*. Vernon (1965) writes that he *"at one time accepted the view that g tends to differentiate into more specialized abilities with age"* (p.29). Guilford (1967), states *"The balance in the evidence seems to be rather decisively against the Garrett hypothesis" (p.58)* More recently, Carroll (1993) writes *" ...there is little evidence to support the hypothesis that cognitive abilities become more and more differentiated with age, up to the period of adulthood... To the extent that differentiation actually occurs, it centres in abilities having to do with different areas of learning and skill formation". (p.687)*

Results which do not support the age differentiation hypothesis have been obtained by a number of researchers. Some studies have suggested increasing ability *integration* with increasing age, whilst others have shown no change or an absence of any consistent trends. Some of these are reviewed below:

Peel and Graham (1948, 1951), gave a battery of ten group tests to pupils at nineteen schools, first at the age of eight and a half to nine and a half, and then two years later. Over the time span, the general factor *increased* in explanatory power from 64.7% to 71.8% for boys, and from 67.2% to 74.9% for girls.

Williams (1948) tested samples of 250 boys at the ages of 12, 13 and 14, using a battery of ten intelligence, spatial and mechanical tests. First factor variances were 51%, 56% and 62% respectively, indicating that *"...secondary education tends to produce greater integration, not specialization of verbal and practical abilities".* (Vernon 1965 p.30).

Chen and Chow (1948) tested samples aged 7-13, and 13-19, and college freshmen, using a battery of ten selected tests. Factor structure became *simpler* with age. A g factor was extracted in all three analyses, with three, two and one additional group factor, respectively, in the three age groups.

Vernon conducted a number of investigations of ability patterning in armed services personnel during the 1940s. In one of these, 1171 boys leaving school at 14 were examined using the standard British Naval battery of five tests, and the results compared to those of 205 seamen recruits who had also left school at 14 in the same district four years previously (Vernon and Parry, 1949). Average inter-correlations and g-saturations were almost identical in the two samples (14 yrs vs 18 yrs). Group factor structure and patterns of correlations differed substantially however, therefore the overall g-saturations may have obscured the true patterns of development. For example, correlations between tests of fluid intelligence (e.g. reasoning, spatial and mechanical tests) and crystallized intelligence (e.g. arithmetic and 'educational' tests) tended to drop.

Swineford (1948, 1949) gave a battery of general, verbal and numerical tests to 952 boys and girls aged between ten and fifteen years. Results indicated no tendency for either a decrease in g factor strength or an increase in the proportion of overall variance explained by group factors. Swineford analysed the scores of her testees by grade, not chronological age, however. At the time of the study it was common school policy to retain poorly performing children in the lower grades. This would have had the effect of confusing chronological age with IQ or Mental Age.

O'Neill (l962) tested six groups from 7.5 to 50 years of age with WAIS and WISC tests and then compared the angles of separation of the two rotated factors. He found factor structure to remain constant over age.

Weiner (l964) tested two separate samples of l400 subjects aged 14 to 54 years using the General Aptitude Test battery. Results revealed little change in the relative contribution of general intelligence, with an apparent slight trend for g to become more important with increasing age.

Wallbrown et al. (1973) reanalysed the matrices of inter-subtest correlations relating to six age groups within the standardisation sample of the Wechsler Preschool and Primary Scale of Intelligence. Wherry-Wherry hierarchical factor analysis revealed a hierarchical factor structure similar to that proposed by Vernon (1950). Factor structure remained stable across age groups, however, contradicting the differentiation hypothesis.

Carroll (1993) reports a reanalysis of fourteen comparisons involving sets of test scores at different childhood ages or school grades.* Some of these were derived from standardization samples and others from experiments designed to test the age differentiation hypothesis. A number of these have been reviewed in this chapter.* Carroll does not give

* [Anderson & Leton,1964; Garrett, Bryan & Perl, 1935; Jones,1949; Paraskevopolous & Kirk, 1969; Reinert et al.,1965; Richards & Nelson, 1939; Reiben & Mengal,1977; Schultz, 1980; Smith & Smith,1966; Sullivan,1973; Sumita & Ichitani,1958; Toussant,1974; Wachs, 1981; Woodcock & Johnson, 1977-78.]

figures for the strength of Spearman's g (e.g. the size of the first factor), reporting only the number of factors extracted with an eigenvalue greater than 1. He observes, however, that *"In nearly all instances, the same number of factors was extracted at each age"* (p.679). Carroll also reanalysed the standardization data for the British Ability Scales (Elliott, 1983), comparing the factor structures derived from the eleven sub-tests at ages 5-7, 8-13 and 14-16. Four factors were revealed at all ages. From these results, he deduces that there is no evidence to support the age differentiation hypothesis.

In a recent Swedish study, Werdelin and Stjernberg (1995) tested all children in school grades 5, 6, 7 and 8 (ages 11 to 15 years) in one district, using a battery of twenty-three ability measures. In total, their sample comprised 1405 subjects, distributed approximately evenly between sexes and school grades. A method developed by Joreskog (1963) was used to extract factors. (The reader is referred to the original paper for a precise description of this method.) This revealed *"six or seven"* factors at grade 5, *"about six"* at grade 6, *"between eight and ten"* at grade 7 and *"eight or nine"* at grade 8. Although rather more factors were found in the older samples, the authors maintain that this provides only weak support for the differentiation hypothesis, since the five main factors were fairly stable over grades. *"There was no tendency for them to split up or differentiate. Also, the correlations between the factors remained almost the same over the four year period"*. They hypothesise that the additional specific factors are a consequence of educational specialisation, rather than a change in the nature of the underlying trait. Nonetheless, the greater number of factors found in the older, as compared to the younger, samples, implies that the influence of Spearman's g has weakened with age. No figures are given to indicate g-strength, however.

Bickley, Keith and Wolfle (1995) compared intercorrelations between the 16 subtests of the Woodcock-Johnson Psychoeducational Battery-Revised (McGrew et al., 1991) in seven age groups drawn from the standardization sample (6, 8, 10, 13, 30-39, 50-59 and 70-79 years). LISREL analysis confirmed that *"neither the correlation matrices among subtests nor the factor structure...differed significantly across age groups"* (p.323)

Deary et al. (1996) compared correlations among the scales of the Differential Aptitude Test in sub-groups of different ages extracted from the Irish normative sample. Two age groups (mean=170 vs 201 months) were compared at each of two levels of mean IQ (90 vs110). Average correlations were not significantly different in the two age groups, irrespective of sample IQ. (This study will be described in more detail in later chapters.)

Guilford (l967) argues that some of the best evidence against the age differentiation hypothesis is the finding of differentiated abilities in very young children, without signs of a g factor. Cited are studies by Kelley (l928), Hurst (l960), McCartin and Meyers (l966), Stott and Ball (l963) and others, all of whom discovered discernible independent abilities at the pre-school level. Carroll (1993) also follows this line, writing *"As far as presently available knowledge indicates, all the major types of cognitive ability are observable, differentiable, and measurable from early in the life span..."* (p.687)

Further explanations for the progressive differentiation of abilities during child development

It will have become clear that while several theorists have observed a progressive differentiation of abilities with age, their explanations for such findings have varied. Some (e.g. Burt, 1954; Spearman, 1925) have emphasised the effects of biological maturation, while others (e.g. Garrett, 1942) draw attention to the effects of learning and experience. There are others still (e.g. Vernon, 1965) who see age-related differentiation as an epiphenomenon of subject and test selection. The purpose of the next section is to present these theories in greater detail.

Effects of experience

Anastasi included a brief review of the age differentiation literature in her book of l964. She concluded that the most likely cause of increasing ability specificity is experience, with like tasks being learned together, and targeted abilities being practiced more. She also speculated that individuals' innate propensities may lead them towards specific interests which might influence the development of patterns of ability. In her words: "*..It may be that the individual's reactional biography tends to shape the very group factors identified by factorial techniques. Such group factors would thus be expected to follow the lines along which educational, vocational, and other life experiences have been organised in a particular culture. Types of tasks that are learned together - in school, on the job, or elsewhere - would as a result tend to become correlated.*" Anastasi cites a study by Fillela (l957) which revealed differences in factor structure between samples from schools having different curricula. Using the same battery of six tests, the two factors which emerged most clearly in technical schooboys were quantitative and spatial-mechanical reasoning. Among academic high school students the clearest factors were verbal and non-verbal reasoning. (These theories are further developed in a paper published by Anastasi in 1970.)

Guilford also drew attention to the role of experience in his book of 1967, citing a study by Mitchell (l956) to support this view. Mitchell compared samples of ll-l2 year olds from high versus low socioeconomic backgrounds, using Thurstone's tests of the five primary mental abilities, plus a few other tests. While essentially the same factors were identified in the two groups, these were more clearly differentiated in the high-SES group. Consistent results were revealed in Sharma and Triptish's (1990) recent study of Indian males and females differing in SES. Intelligence was more differentiated in samples of higher, compared to lower SES and in boys, compared to girls (it could be argued that, in India, boys have access to a more diverse range of educational opportunities).

Carroll (1993) also interprets the age differentiation literature in this way, maintaining that it is mainly factors relating to *"specialised learnings"* that differentiate, rather than those relating to basic cognitive abilities. Likewise, R.B. Cattell (1971) hypothesised that the structure of crystallised (rather than fluid) intelligence depends on educational profile. Such a theory might explain Vernon and Parry's (l949) finding of lower correlations between fluid and crystallised intelligence in older as compared to younger teenagers.

Selection of tests and subjects

Vernon (1965) sees subject and test selection as the chief explanatory variables. He writes *"so much depends on the degree of heterogeneity of the people (and tests) concerned".* As noted earlier, specific factors may account for a greater amount of overall ability variance in highly selected groups than in those which are more heterogeneous in terms of ability. This, claims Vernon, is why the g factor is weaker in college groups, not because they are older or brighter.

While the effects of subject selection are undoubtedly important, they have been controlled for in many of the studies cited and cannot be considered wholly causal to the findings. The question of test heterogeneity is less easily reconciled, however. Findings of greater differentiation of abilities with age may be an artifact of the types of tests used. For example, at very young ages few tests can be administered, since infants cannot read or write, hence elementary reactions to stimuli, motor control and other, very general, skills are all that can be assessed. As the child becomes older, both ability to respond and learned skills will increase in number and diversity. Specific abilities will gradually become discernible from the overall or general ability, and factor structure will change accordingly. Garrett (1942) also acknowledges the role of test selection. He notes that the child who talks and reads earlier will surpass his age-mates in all other subjects because at this age most tasks are linguistically orientated. However, *"with increasing maturity and with a more nearly common background of language facility, general ability dissolves into more specialised talents or group factors"* (p.376). Carroll (1993) and Brody (1992) also draw attention to the role of test selection in determining factor structure.

Could developmental differences in general ability underlie the age differentiation phenomenon?

The results of many longitudinal studies indicate that performance on tests of general intelligence and specific abilities increases steadily during childhood. There is some debate as to whether these performance increments are a result of improvements in overall intellectual capacity or whether they are a product of knowledge aquisition. If such changes do, to some extent, reflect a growth in fluid intelligence, it might be speculated that it is this increase which gives rise to differentiation. This would tie-in with the hypothesis that there will be a greater degree of ability differentiation in persons of high versus low intelligence. The theory that it is intelligence level, rather than age *per se*, which underlies the progressive differentiation of abilities with age, is not without advocates. As was noted earlier, Burt (1947) writes that "...with the younger *or more defective*, we can measure manifestations on the lower levels only. With the older *and brighter* we shall need to measure higher mental processes" (p.77), implying that general intellectual level improves during childhood. This was the essence of Spearman's (1925) explanation for the differentiation of abilities during childhood, described earlier. German psychologists Leinert and Crott (1964) and Rienert (1965) also theorised along these lines.

The results of several studies using psychophysical measures, such as Inspection Time (IT), offer support for the theory that fluid intelligence increases during child development. Inspection Time has been hypothesised to tap fluid intelligence by accessing mental speed or neural efficiency. Faster IT is associated with higher scores on intelligence tests. Nettelbeck (1987) reviews several studies which have indicated an increase in speed of input processing with ontogenetic development. Nettelbeck compares the results of studies by Hulme & Turnbull (1983) and Smith & Stanley (1983) which used similar procedures. Mean IT among Hulme & Turnbull's 6-7 year-olds was significantly longer than that for Smith & Stanley's 12-13 year-olds ($z = 4.72$). Furthermore, reanalysis of Smith & Stanley's data revealed significantly longer ITs among 12 than 13 year-olds ($p<.01$). Direct evidence that IT becomes faster with ontogenetic development, at least until age 11 to 13 years of age, is provided by a series of investigations by Wilson (1984). Wilson employed a cross-sectional design to compare IT among children aged 6 to 12 years and adults. A longitudinal follow-up was conducted on most children, so as to distinguish maturation effects on IT from effects due to cohort and practice. Longitudinal change after one year could not be explained as resulting from practice, and cross-sequential analysis established that IT became shorter with increasing age, independently of cohort.

The problem in these studies was the confounding of mental age (MA) and chronological age (CA). While the two are strongly related in homogeneous groups of average intelligence, this is not the case when (as usually happens) groups are heterogeneous with respect to IQ. To disentangle the effects of CA and MA, Wilson compared children at different IQ levels (Raven's Coloured Progressive Matrices) while controlling for MA. IT was measured for two MA groups (8 and 11 years), each consisting of below average but nonretarded (IQ 80-90), average (IQ 95-114), and above average (IQ 122-135) subsets ($N = 8$). Consistent with other experiments in her thesis, Wilson found significantly slower IT in the MA-8 group. Within MA groups, however, IT did not correlate significantly with IQ. This finding suggested that, if retarded persons are excluded, it is MA (absolute ability level) that correlates with IT until adult levels of performance are reached.

Findings of developmental improvements in IT performance lend credence to the theory that maturation is associated with a change in some central processing mechanism (e.g. Brand, 1981, 1984; Kail,1991, 1992).

The hypothesis that increases in fluid intelligence underlie age differences in the strength of g, is supported by findings indicating increasing *integration* of abilities from early to later adulthood. Scores on tests of fluid intelligence and performance on tests of mental speed tend to decline with aging, just as they improve during childhood (e.g. Nettelbeck, 1987).

As described earlier, Balinsky (l94l) found a progressive decrease in the variance attributable to g in samples aged 9, 12, 15 and 25-29 (38, 36, 24, 20 percent, respectively). In samples aged 35-4 and 50-59 years, however, g accounted for 33 and 45 percent of the variance, respectively. Similarly, while Lienert & Crott (1964) reported a decrease in the proportion of variance accounted for by the first centroid factor, from 48% at age 10-12 to 45% at age 18-20, g explained 47% of the variance in a sample aged 45-60.

Stricker and Rock (1987) report complementary results for three samples aged 20-29, 30-39 and 40-49 years. Confirmatory factor analysis of the items on the Graduate Record Examinations (GRE) General Test revealed the same ability factors in each sample (Quantitative, Verbal, Analytical), but intercorrelations between the factors increased (slightly, but consistently) with age.

Given these findings, it could be speculated that differences in factor structure associated with age have the same source as those associated with differences in IQ level. Wilson's (1985) findings of a stronger relationship of IT to mental than to chronological age would appear to support such a suggestion.

Theories which posit a relationship between the strength of Spearman's g and level of general intelligence will now be described.

CHAPTER 3

INTELLIGENCE LEVEL AND SPEARMAN'S G

Theories predicting ability differentiation at high IQ levels

Charles Spearman and the Law of Diminishing Returns

Charles Spearman first suggested a theory compatible with the differentiation hypothesis in 1925. This was described earlier, in connection with the differentiation of abilities with age. The subtheory is more particularly concerned, however, with differences in the power of positive manifold across the range of general ability.

To recap, Spearman (1925) likened the mind to a factory composed of several engines which are specialised for different tasks, and a central energy source which powers them. He hypothesised that individual differences in levels of energy (associated with global neurological efficiency) give rise to g and that specific factor variance is attributable to the specific engines. The amount of energy available predicts how many specific engines can be used at any one time and the efficiency with which they are able to function. The relative importance of the energy and the engines varies depending on the nature of the task. Complex tasks involve many engines simultaneously, hence general energy (e or g) is the best predictor of performance on such tasks, while the engines (specific abilties) are more important for those which are simple. Similarly, the energy is more important for tasks which are novel and the engines for those which are well learned.

More importantly, with respect to the differentiation hypothesis, is Spearman's prediction that *"the relative influence of the energy and the engines changes largely with the class of person at issue"*. By "class" Spearman is referring to level of basic ability (g or e). He theorised that the higher an individual's level of basic mental energy (or IQ), the less likely it is that their performance on cognitive tasks will be predicted by g, and the more important will be specific abilities. To explain this, he draws on the theory of the 'law of diminishing returns' :

> "...the more energy available already, the less advantage accrues from further increments of it. And this is a well known property of engines in general. Suppose that a ship at moderate expenditure of coal goes 15 knots an hour. By additional coal the rate can be readily increased another 5 knots. By doubling the addition of coal, however, the additional knots will certainly not be anything like doubled." (1925, p.7)

According to Spearman, the law of diminishing returns applies to intelligence and leads to the prediction that *"the influence of g is lower in just the class of person who possess it most abundantly"*.
Put simply, in persons of low intelligence the level and complexity of the tasks which can be performed is constrained by the amount of mental energy available. Because the capacity to perform most tasks is very dependent on e-level in such individuals, scores on tests of specific abilities will be highly correlated, giving rise to a strong g factor. Those with a surplus of energy (high e or g), however, are less dependent on it because their engines are already able to function to their maximum potential. Since the engines vary in quality, their potential performance is constrained primarily by their own limitations. The individuality of different ability traits is therefore more obvious (and the factorial structure more divergent) in those with higher, versus lower, intelligence.

The ideas expressed in Spearman's paper of 1925 are compatible with many modern theories. For example, his thesis of general mental energy has been supported by investigators of the biological basis of intelligence (e.g. Weiss, 1986). There is also evidence to support the idea that complex tasks are more dependent on g than simple tasks (eg. Jensen, 1982). Spearman's distinction between learned and novel tasks, with respect to their reliance on g, also anticipates the distinction between crystallised and fluid intelligence proposed by Cattell and Horn (l966) and between *automatic* and *controlled* processing, as postulated by cognitive theorists such as Norman & Bobrow (1975). Several researchers have reported that performance is less dependent on g for tasks which are novel than those which have been over-learned or automatised (eg. Norman & Bobrow,1975; Schneider & Schiffrin, 1973; Ackerman,1986). Ackerman (1986), for example, presented subjects with sets of tasks which varied in stimulus consistency, such that practice in one set could lead to the development of automaticity whereas practice had little effect on the other set. He found that task performance was always g-dependent for the novel or inconsistent tasks, whereas in tasks containing consistent characteristics, performance was at first g-dependent but gradually became g-independent as automaticity was achieved. The speed with which automaticity was reached was largely predicted by g and broad content abilities, as would be expected if acquisition depends on central resources. Ackerman also notes that not all subjects will be able to achieve automaticity in certain tasks since they differ in the amount of available *"total controlled processing resources* [e or g]". He speculates that *" Task difficulty level may act as a high pass filter so that only subjects of a certain ability level can be successful"*. This would accord with the differentiation hypothesis in that subjects of higher g will be able to acquire and use a greater number of skills.

The German School: The theories of Wewetzer, Lienert, and Rienert

German psychologist, K. Wewetzer, first put forward the hypothesis that intelligence will be more differentiated in groups of higher intelligence, in 1957. He referred to this as the *'divergence hypothesis'* to distinguish it from Garrett's hypothesis regarding age differences in g-strength.

G.A. Lienert (1959, 1960) supported Wewetzer's theory of divergence, developing it in such a way as to integrate it with Garrett's theory of differentiation with increasing age. Lienert and Crott (1964) write:

> "In the same way as the improvement in performance from childhood to adolescence is accompanied by a differentiation of the underlying ability structure, and as the decline of performance from adolescence to adulthood is accompanied by an integration of structure, a divergence and a convergence of the ability structure takes place from low to high performance level and vice versa respectively. Therefore, it may be assumed that performance level and degree of differentiation are interdependent, and that - regardless of the subject's age - differences in intelligence level presuppose variations in the structure of intelligence in accordance with the divergence hypothesis." (p.158)

Reinert (1964), took on board Wewetzer's theory, later labelling it the "Leistungsdifferenzierungshypothese" or "performance differentiation hypothesis" (Reinert, Baltes & Schmidt, 1965). Reinert agreed with Lienert's view, arguing that the extent to which abilities are differentiated is related to the *absolute ability level* of the sample, rather than age *per se*.

The findings of these theorists will be reviewed in the next section.

Anderson's Model of Minimum Cognitive Architecture (or Theory of Multiple and General Intelligences)

The modern theory of intelligence which is of greatest relevance to the differentiation hypothesis, is that proposed by Michael Anderson. Indeed it was an analysis of this model which stimulated the present investigation. First suggested in 1986, Anderson's model has been refined and modified in a series of subsequent papers (Anderson, 1986; 1987; 1989a; 1989b) culminating in a book published in 1992.

Although Spearman's theory was unknown to Anderson at the time he proposed his version, the two bear remarkable similarities. Nonetheless, Anderson's theory has several distict features.

The model accounts for both hypothetically independent abilities and a pervasive g factor. It also posits a relationship between general, group and specific factors which gives rise to the prediction that Spearman's g will be weaker at higher levels of general intelligence.

Anderson extends the issue of differentiation beyond the psychometric domain by using concepts from information-processing theory. Although a number of models of intelligence have emanated from this field (eg. Sternberg,1983,1984,1988; Hunt,1975, 1978, 1980) and from the newer area of cognitive science (e.g. Fodor,1983; Gardner,1983), their insistence on examining vertical rather than horizontal processes (looking not at ability differences between individuals but at very specific ability processes within subjects) has constrained their usefulness in terms of describing population parameters. Anderson's theory is valuable because it accounts for both structural components of intelligence and individual differences.

Anderson (1986) begins by outlining the well known psychometric evidence in favour of Spearman's g.

> "Attempts to measure hypothetically independent abilities have been unsuccessful; they always covary. This is in marked contrast to the successful attempts to design tests which measure g and very little else...Multiple independent factors would predict that differences between abilities should be as great within individuals as between individuals. The fact that g can be found at all indicates that this is not so"(p. 298)

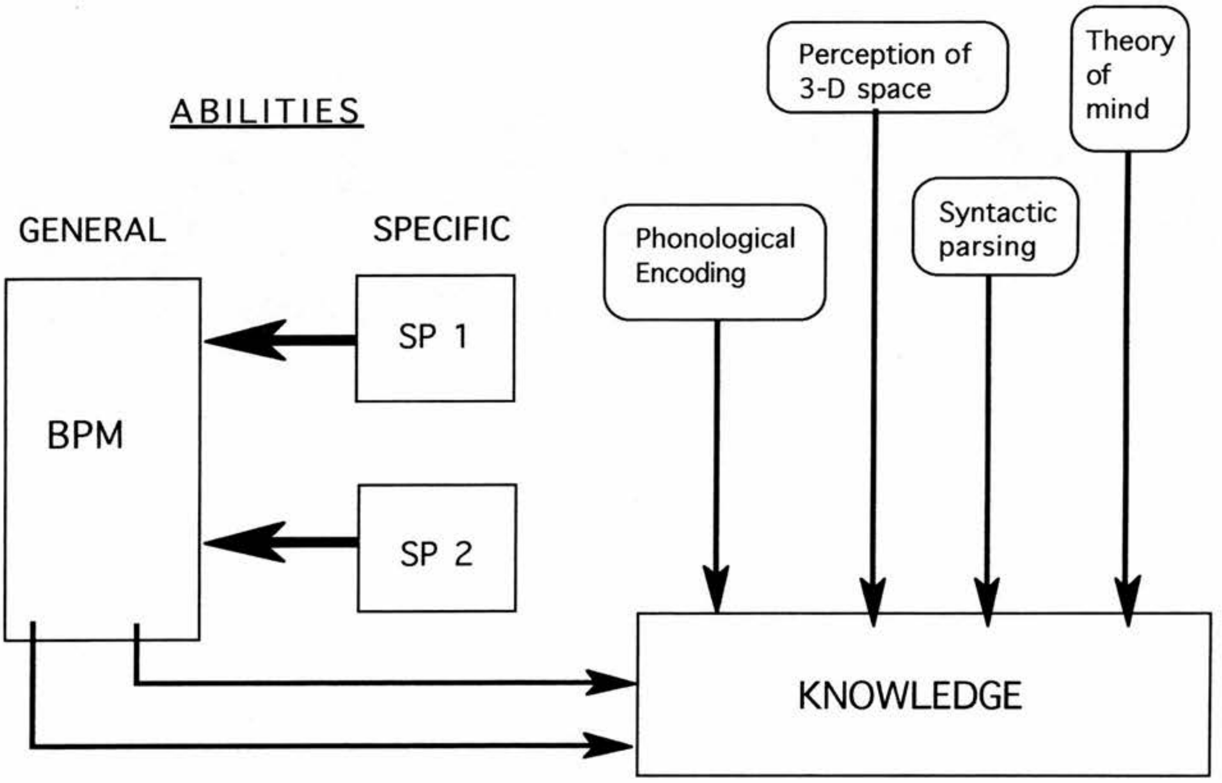
Although he believes g to be of overriding importance in explaining intelligence, he acknowledges that it does not fully account for inter-test variance.

> "Not all variations in ability are general....Psychometrics, cognitive psychology, genetics and neurophysiology all propose that there are also individual differences in more specific abilities" (Anderson,1989b. p.19)

Anderson draws attention to exceptions to the law of positive manifold, pointing out that there are certain competencies, such as the ability to perceive a stable three-dimensional world, which are essentially invariant within a population and appear to be independent of g. Although these are not normally considered 'intellectual', they do contribute to overall ability and, as such, must be incorporated into a comprehensive theory of intelligence. The phenomenon of the 'idiot savant' also illustrates the fact that certain 'abilities' apparently function independently of g. Such individuals are characterised by a general cognitive deficit, but exhibit one or more normal or outstanding abilities (e.g. in art, music, or mental arithmetic). Conversely, some children and adults of high general intelligence show marked weaknesses in specific domains such as pitch perception or mechanical aptitude.

According to Anderson, general intelligence, specific abilities and independent skills can all be accounted for within one theoretical framework. Figure 3.1 (overleaf) is said to illustrate the minimal cognitive architecture necessary to explain the variation of intellectual abilities within and between individuals.

Figure 3.1 Cognitive Architecture Hypothesised by Anderson to Underlie Human Abilities (adapted from Anderson,1989b)



According to Anderson, the devices represented on the upper right hand side of figure 3.1 are independent, task-dedicated "**modules**", similar to those postulated by Fodor (1983). These *"do not vary between individuals in the quality and/or quantity of their output...They ... either work well or not at all"* (Anderson, 1986). The examples shown here represent just a few of the apparently complex skills which are invariant within a population.

The devices on the left hand side of the figure, in contrast, do show individual differences. The Specific Processors (SP1 and SP2) are described as *"computational devices that generate algorithms for solving cognitive problems".* Anderson likens them to two computer programming languages (e.g. Fortran and Pascal), each of which is more suited to dealing with particular kinds of problems, but both of which are, in principle, capable of solving any problem. Anderson regards one as a spatial/analogical processor and the other as a verbal/ propositional processor. This distinction is consistent with the broad group factors of visuospatial and verbal ability identifed in hierarchical models of intelligence and with functional differences between the right and left cerebral hemispheres (e.g. Corballis, 1989).

The basic processing mechanism (BPM) is the most important part of the system. Anderson describes it as *"the mechanism which instantiates the algorithms generated by the specific processors"*, or the computer hardware on which the software runs. The solutions it produces become "knowledge" which is either actively executed or stored for later use. In contrast, the "modules" give rise to skills or knowledge without the assistance of the BPM. They can be thought of as 'read only' programs in a computer, which perform tasks directly without having to take up space in the active buffer space of the computer's RAM (random access memory - analogous to attention).

The basic processing mechanism varies between individuals in its efficiency, or e. Anderson identifies e with speed of information processing, citing work on inspection time and reaction time to support this contention. He also refers to the fact that correlations between g and knowledge-dependent IQ tests are often no higher than those between g and knowledge-independent psychophysical measures.

At one level SP1 and SP2 are independent: The power of one does not predict the power of the other and neither predicts the power of the basic processing mechanism. However, the efficiency of the basic processing mechanism constrains the implementation of the algorithms generated by the specific processors, just as Spearman's "energy" constrains the performance of his "engines". The 'abilities' produced by the two specific processors thus become correlated, so it is individual differences in BPM efficiency which are responsible for g. It is important to note that

> "e is not g and computational devices [modules and SPs] are not specific abilities. They are at different levels of abstraction. g and specific abilities are abstractions from population performance on cognitive tasks, e and computational devices are abstractions from the internal cognitive architecture of an individual" (Anderson, 1986 p. 304)

The major prediction to arise from Anderson's model is that the influence of Spearman's g will decrease as general ability level increases. He explains this as follows:

> " At low e there will only be poor output from both specific processors, irrespective of their latent power. At higher e differences between the power of the different processors will become more apparent and influential. Thus, because the code generated by the specific processors has to be implemented on a cognitive machine, the basic processing mechanism, and this mechanism varies in its efficiency, e, it is e that gives rise to general intelligence from what is a collection of different cognitive mechanisms" (Anderson, 1987 p. 5)

In short, because the BPM constrains the output of the specific processors, it is only at higher e levels that their potential can be expressed. Low e levels, in contrast, are associated with a higher degree of positive manifold. It can also be speculated that certain problem -solving algorithms will not function *at all* at lower e levels if they need a certain e 'threshold' in order to operate. Furthermore, low e might be expected to inhibit the use of algorithms which are dependent upon the efficient utilisation of a preceding algorithm in the e-dependent hierarchy.

Anderson's model bears similaries to the 'working memory' model of Baddeley and Hitch (1976), who postulate three mechanisms operating in short term memory: a modality-free 'central executive'; a 'visuospatial scratch pad' and an 'articulatory loop'. The most important component is the central executive, which resembles attention. It has limited capacity and is thought to be used when dealing with most cognitively demanding tasks. The articulatory loop and visuo-spatial scratch pad are 'slave systems' that can be used by the central executive for specific purposes (eg. verbal or spatial coding). The distinction between SP1 and SP2 is also compatible with the different modes of thought hypothesised by many information-processing theorists. The distinction between analogical and propositional forms of information processing has been made by various cognitive scientists. For example, Kosslyn (1980,1981,1983) sees knowledge as being stored in coded, meaning-based, abstract units. These units are language-like representations which assert facts about the world and can be thought of as concepts or ideas each of which has a number of associations. They can be likened to the bits and bytes of a computer's data base. Kosslyn hypothesises that there are two ways of acquiring and of using such knowledge. In the first, information is taken into or out of the knowledge store in a series of propositions. This would be expected when the information of interest is predominantly verbal or involves relating assertions and facts. In many cases however, propositions are inadequate for knowledge generation, such as in tasks involving visuospatial reasoning. In such cases the information may be handled by means of directly representational images. (The propositional/analogical distinction may also be related to differences between declarative and procedural knowledge, respectively.) Although Anderson does not explicitly refer to these referents of SP1 and SP2, it is obvious from his approach, that he has considered much diverse literature in developing his model.

Anderson's theory is remarkably similar to Spearman's, although it has many unique features, such as the inclusion of invariant modular abilities.

The two theories differ very strongly, however, in regard to the issue of *age differences* in ability structure. In view of the evidence for increases in measured intelligence with age, especially increased scores on tests of fluid intelligence, one might expect Anderson's developmental position to parallel that regarding individual differences. Paradoxically, however, this is not the case. Anderson (1989) writes: *"The major new idea suggested by this theory is that intelligence does not develop"* (p.93). By 'intelligence' Anderson is referring to the e-level of the basic processing mechanism, which is held to be fixed at birth, rather like Hebb's intelligence A. According to Anderson:

> "The hypothesis that the speed of the basic processing mechanism is unchanging through development explains why individual differences are so stable, why it is that time-based measures in infancy research are the best predictors of later individual differences, and why children of different levels of ability seem to follow different developmental profiles." (1989, p. 93)

Cited as evidence for stability in individual differences throughout development, are studies showing that infant habituation time in infancy is a good predictor of IQ differences in later childhood (e.g. Fagan & McGrath, 1981; Rose, Slater and Perry, 1986).*
Anderson does accept that performance on intelligence tests improves during child development, but he attributes this to a) the fact that the older child will have accumulated more knowledge which s/he can apply to the solution of new problems and b) the coming on line, or maturation, of the various modular capabilities, which will lead to *"across-the-board increases"* in ability. He refers to his 1988 study of the relationship of age to reaction time and inspection time. Age was found to be only weakly

*A recent meta-analysis by McCall and Carriger (1993) confirms such findings. Furthermore, these authors report that the predictive validity of infant habituation measures for later IQ performance is greatest in 'at risk' (including low IQ) groups, a finding which is consistent with the ability-differentiation hypothesis.

correlated with IT but strongly related to RT. This had been predicted on the basis of the theory that RT is more prone to knowledge and strategy influences than IT, whereas the latter is a purer indicator of innate processing speed. Anderson believes, therefore, that it is growth in knowledge and the maturation of modular capacities which is responsible for increased general ability during childhood.

That the theories of Anderson and Spearman should be so alike, is not as surprising as it might, at first, seem. The ideas expressed in the two models have an intuitive ring and common features can be detected in many other theories explaining mental performance, such as the working memory model described earlier. Indeed, Douglas Detterman (1987) had proposed a similar theory in a paper concerning the nature of mental retardation, to which he made reference when attempting to explain his later findings of a stronger g factor in groups of lower IQ. A brief account of this theory is given by Detterman and Daniel (1989, p.358):

> "Intelligence is a system made up of a small number of independent processes. Mental retardation is caused by deficits in central processes, meaning processes which most heavily affect all other processes in the system. Because of the deficit in the central processes the entire system is brought to a uniform low level of operation. So all processes in subjects with deficits tend to operate at the same uniform level. However, subjects without deficits show much more variability across processes because they do not have deficits in important central processes. This causes high correlations among mental measures in low IQ subjects and low correlations in high IQ subjects"

At the time this was written, Detterman was unfamiliar with the theories of differentiation put forward by Spearman and Anderson. Such similarities across theories indicate that the fundamental hypothesis is an important one which warrants further investigation. Findings of a consistent relationship between ability level and the magnitude of positive manifold would have significant implications for psychometric theory. A substantiated theory of this type would also be valuable from the point of view of reconciling monistic and pluralistic viewpoints.

Intelligence level and the differentiation of abilities: A review of the evidence

Despite their potential value for theoretical and applied psychology, theories of differentiation have been surprisingly neglected over the last few years. A striking illustration of this is the following recent statement by Douglas Detterman, editor of the flagship journal *Intelligence:*

> "During the 85 year history of this work, it was thought that the positive manifold was uniformly distributed over the full range of ability. That is, it was assumed that the correlation among mental tests would be about the same in a group of low IQ subjects as it would be in a group of high IQ subjects." (Detterman & Daniel, 1989 pp. 349-350)

As detailed in the last chapter, the issue has, historically, been an important one and that it should have been neglected by a generation of researchers is intriguing. This writer and her supervisors were as ignorant as the latter theorists when the research for this thesis began, at which time Anderson appeared to have been the first to have specified the relationship between IQ level and strength of g.

While the work of Spearman and others may have been overlooked, many writers have, over the years, alluded to the idea that the effect of positive manifold may differ across the IQ spectrum. Vernon (1965), for example, declared that *"it (g) corresponds pretty closely to what we mean by 'intelligence', provided we admit - with Guilford - that this differentiates into a wide range of abilities at higher levels"* (p.139) Similarly, O'Connor & Hermelin (1983) stated, *"It is only at higher IQs...that heterogeneity in cognitive abilities assumes any importance in accounting for individual differences".* (p.395)

In view of the widespread lack of knowledge amongst intelligence researchers, regarding theories of differentiation, there has been little research which has specifically addressed the issue. Although this situation has improved since the publication of Detterman and Daniel's (1989) paper, it is essential to extend the net widely when searching for evidence on this topic. Many investigators have reported incidental findings of differentiation in studies with other aims. The creativity literature, for example, is a very useful source.
Let us now turn to the review:

In the paper in which he outlines his sub-theory of the law of diminishing returns, Spearman (1925) refers to evidence from a study by Abelson. Abelson tested groups of normal and mentally defective children using the same battery of twelve tests. The inter-test correlations are shown in tables 3.1 (a) and (b). As can be seen from these tables, test scores inter-correlated to a lesser extent in the sample of normal children.

Table 3.1 (a) Correlations between mental tests in 78 normal children tested by Abelson. Reported in Spearman (1925)

	1	2	3	4	5	6	7	8	9	10	11	12
1) opposites												
2) observation	75											
3) absurdities	78	72										
4) memory sentences	71	58	53									
5) crossing O's	62	60	41	54								
6) geometrical figs.	64	58	44	61	73							
7) discrim. length	72	67	79	54	48	45						
8) crossing patterns	78	56	68	37	54	48	56					
9) memory form	57	58	41	54	38	30	49	30				
10) tapping	40	56	46	55	36	42	30	21	24			
11) grip strength	46	52	34	19	52	48	31	27	29	20		
12) interpret. pictures	33	29	29	43	35	35	06	18	28	18	28	

Table 3.1(b) Correlations between mental tests in 22 'defective' children tested by Abelson. Reported in Spearman (1925)

	1	2	3	4	5	6	7	8	9	10	11	12
1) absurdities												
2) opposites	100											
3) crossing patterns	100	97										
4) crossing O's	98	95	91									
5) memory sentences	97	87	80	85								
6) observation	100	91	88	77	73							
7) memory form	100	85	68	84	90	76						
8) interpret. picts.	100	76	92	67	68	83	65					
9) geometrical figs.	98	85	74	76	88	71	67	74				
10) discrim. length	94	87	78	81	65	86	70	80	65			
11) tapping	94	70	76	73	78	59	77	80	60	51		
12) grip strength	79	72	67	55	68	65	75	59	62	45	61	

Abelson's data were subjected to principal components analysis by Deary and Pagliari (1991), who reintroduced Spearman's theory to the intelligence community, following Detterman and Daniel's faux pas. The results of this analysis are shown below:

Table 3.2 Principal components analysis of Abelson's data reported by Spearman (Deary & Pagliari, 1991)

	NORMALS		DEFECTIVES
FACTOR	1	2	1
Percentage of total variance	52.6	10.1	80.8
TEST			
opposites	.92	-.17	.97
observation	.87	-.06	.90
absurdities	.81	-.36	1.00
memory for sentences	.77	.15	.91
crossing out Os	.76	.28	.92
geometrical figures	.75	.31	.86
length discrimination	.77	-.46	.84
crossing out patterns	.71	-.36	.94
memory for forms	.62	-.01	.88
tapping	.57	.19	.82
grip strength	.56	.34	.75
interpreting pictures	.43	.60	.87

Two factors with eigenvalues exceeding 1.0 were extracted in the normal sample and only one in the sample of defectives. The proportions of variance accounted for by the first principal component were 52.6% for normals and 80.8% for defectives.

Results consistent with the differentiation hypothesis were obtained by Vernon in a study of U.S. army and navy personnel tested in 1947. Vernon (1965) reports:

> "When the same tests which, among unselected recruits gave g and group factor variances of 50 percent and 20-25 percent respectively, were analysed among high-grade mechanics and officer candidates, g often fell to 15 percent and group factors rose to 35 percent" (p. 360)

Vernon attributed these findings to restriction of range in the higher ranking sample. However, given the large numbers of such subjects in the 1947 cohort, the distribution of ability amongst higher-grade personnel is likely to have been wide, hence selectivity effects may not have been entirely to blame.

Other studies of US military personnel, conducted around the same time as Vernon's work, yielded similar results. Correlations between tests of various verbal, numerical, and spatial abilities; whilst low and negligible in college students; were high and significant in forces recruits (U.S.Army TAGO,1945; U.S. Naval test battery,1945). Garret (1946) highlights the strong g factor arising from the analysis of the scores of some 8,000,000 men on the Army General Classification Test. He writes:

> "This probably resulted from the fact that many soldiers were undoubtedly closer to the elementary school child than to the superior adult in the facility with which they handled abstract test material." (p.377)

Findings by Wiseman (1964) also support the hypothesis. Wiseman administered a number of specially constructed tests of intelligence, reading comprehension and mechanical arithmetic to all fourteen year olds in Manchester in 1951. Also included were measures of a number of 'social' variables. From these subjects a 'bright' and a 'backward' group were selected with average intellectual scores over one standard deviation above or below the mean, respectively. Inter-test correlations were substantially higher in the lower ability group, indicating a strong g factor, while in the bright group the correlations tended to be low or negligible.

It is important to note that greater ability differentiation in more intelligent groups does not imply that all high IQ subjects exhibit the same ability patterns. By high differentiation is meant not only a decrease in the strength of the g factor, relative to that of group or specific factors, but an increase in the *variability* of the latter. This would be expected if the overriding influence of whatever is responsible for g, is lessened in the higher capacity subject. This high score/high variability relationship has been noted by a number of researchers (eg. Humphreys, 1982; Ackerman,1986; Spitz,1986). The predictive validity of tests of general intelligence may thus be lower in higher ability groups. Hudson (1966) points out that tests differentiate quite poorly between out-standingly able and more mediocre boys in high grade secondary schools or among university students. In a recent study of 'gifted' school children, Tyler-Wood and Carri (1991) found substantial variation among four estimates of cognitive ability, concluding that multiple criteria are needed for identifying such individuals. Spitz (1986) notes that the reliability of tests is lower amongst subjects of high ability and higher at the lower end of the IQ distribution. Spitz (1982) writes:

> "If a greater amount of the variance is reflected in the g factor in retarded individuals, we would also expect them to show a greater IQ constancy. In fact, there can be little doubt that this is the case. Goodman and Cameron (1978) point out that the correlations on individual tests between 1 year and three years of age are .40 to .50 for nonretarded children, and .70 to .80 for retarded children. Additionally, the lower the IQ the higher the correlation. The predictive value of IQ tests for retarded children is very good indeed ".(p.180)

Wewetzer (1957) compared correlations between Wechsler sub-tests in age-matched samples having low, intermediate and high levels of general intelligence. Results indicated a weakening of Spearman's g with increasing IQ, supporting his hypothesis of 'divergence'.

Bouyer and Kniep (1981) replicated Wewetzer's study in France, using 473 nine year olds, representing all children of that age in one school district. All subjects had been tested with the W.I.S.C. battery. The sample was divided into three groups, having different mean IQs but equal ranges. [Retarded: 49-84 , N=77; Normal: 85-115, N=307; Gifted: 116-139, N=89.] Inter-subtest correlations were subjected to principal components analysis with a varimax rotation and the scree test was used to decide the number of factors to extract. Although three factors were extracted in all groups the proportion of variance accounted for by the first factor declined as IQ increased. The relevant percentages are as follows: Retarded 29.53; Normal 28.33; Gifted 25.19. Bouyer and Kniep interpret their results as supporting Wewetzer's divergence hypothesis.

Lienert & Faber (1963) compared correlations between scores on the scales of the German HAWIK test battery (a version of the Wechsler tests) in samples of nine and twelve year old children with mean IQs above 109 and below 91 (N=150 in all four sub-groups). In both age groups, the first centroid factor was smaller in the sample having higher IQ, indicating a weaker g factor. Comparing high and low IQ groups, the proportions of variance accounted for by the first factor were, respectively, 54% versus 49% in nine year olds and 63% versus 54% in twelve year olds.

Reinert, Baltes and Schmidt (1965) devised an ingenious experiment to examine the hypothesis that it is *absolute ability level*, rather than age or IQ, which underlies the differentiation effect. They selected two groups, from a larger sample, who were equal in "level of performance" (average test score) but different in age and in IQ. They hypothesed that there would be no differences between the groups in factor structure. A second comparison was made between groups differing in absolute ability (high versus low) but equal in IQ. In this comparison, they predicted that the group with higher test scores would show a more differentiated factor structure. Both hypotheses were supported. Although the numbers of factors extracted in each group was the same, the strength of Spearman's g - indicated by the size of the mean correlations and communalities - was equal in the score-matched groups despite their different ages and IQs, and g was stronger in the IQ-matched samples with lower levels of absolute ability.

Anderson (1987) derived evidence for his differentiation hypothesis from a reanalysis of data from the Child Health and Education Study (1980). This is a longitudinal study of around 15,000 British Children which began in 1970 and which aims to test the same cohort every five years. Anderson's analysis is based on data collected in 1975. Children were assessed using an 'educational pack' containing a teacher's rating scale, and a questionnaire dealing with demographic information, and a 'psychological pack' consisting of seven ability tests. Four of these tests were drawn from the British Ability Scales: word definition; matrices; similarities and digit recall. The remaining seven tests were pictorial language comprehension, 'friendly maths', and a shortened version of the Edinburgh Reading Test.

Subjects with IQs below 70 were excluded from the analysis, since this group tends to include those with pathologically induced psychological difficulties. This, and absences, reduced the sample size to 12,905.

In the total sample, matrices of inter-test correlations yielded a single general factor when subjected to principal components analysis. When subjects were divided into two groups, one scoring above and one below the mean on composite IQ, however, the following results were obtained:

Table 3.3 Proportions of variance accounted for by g (F1) and other test variables (F2) in groups with high and low IQ (Anderson, 1987).

TEST	Total population	High IQ group			Low IQ group	
	F1	F1	F2	F3	F1	F2
1) Word Definition	.82	.81	.15	.00	.76	.06
2) Matrices	.73	-.09	.85	-.17	.18	.62
3) Similarities	.80	.72	.06	-.11	.70	.17
4) Digit Recall	.52	-.05	.01	.97	-.28	.65
5) pictorial language comp.	.70	.70	.15	.05	.70	.09
6) Friendly maths	.86	.38	.72	.16	.44	.71
7) Edinburgh reading test	.86	.49	.63	.16	.50	.67

Anderson (1987) writes of these findings:

> "That the factor structure of intelligence should change when the range of general intelligence is restricted in this fashion is unsurprising. However, the prediction... is that the factor structure of intelligence will be different at different levels of general intelligence, and particularly there will be more factors the higher the level of general ability. " (p. 10)

Clearly, this prediction was supported.

As mentioned already, it was a study by Detterman and Daniel (1989), published in the international journal *Intelligence,* which reactivated interest in the differentiation hypothesis. Detterman and Daniel were unaware of the theories of Spearman, Wewetzer, Anderson and others who had suggested that the influence of positive manifold might be lower in samples of high versus low intelligence. They conducted a number of studies to test this hypothesis.

In the first study, intercorrelations of nine basic cognitive tasks with each other and with the 11 sub-tests of the WAIS-R, were compared in a sample of 20 mentally retarded young adults and 20 college students. The groups had approximately the same distribution of IQ but means some 50 points apart. (SDs = 7.56 and 7.79, Means = 67.5 and 115.5, respectively.)

The experiment was repeated (this time using ten cognitive tasks) in two groups of randomly selected high school students having low versus high IQ. (LOW: Mean = 93.0, SD=12.3, N=68; HIGH: Mean=122.0, SD = 9.9, N=73)

Detterman and Daniel report the following results for these two studies:

Table 3.4 Average correlation of basic cognitive task measures with IQ scores on the WAIS-R and with each other (Cognitive) (Detterman and Daniel, 1989)

	Retarded & College Adults		High School Students	
	IQ	Cognitive	IQ	Cognitive
IQ level				
Low	.60	.44	.37	.26
High	.26	.23	.24	.18

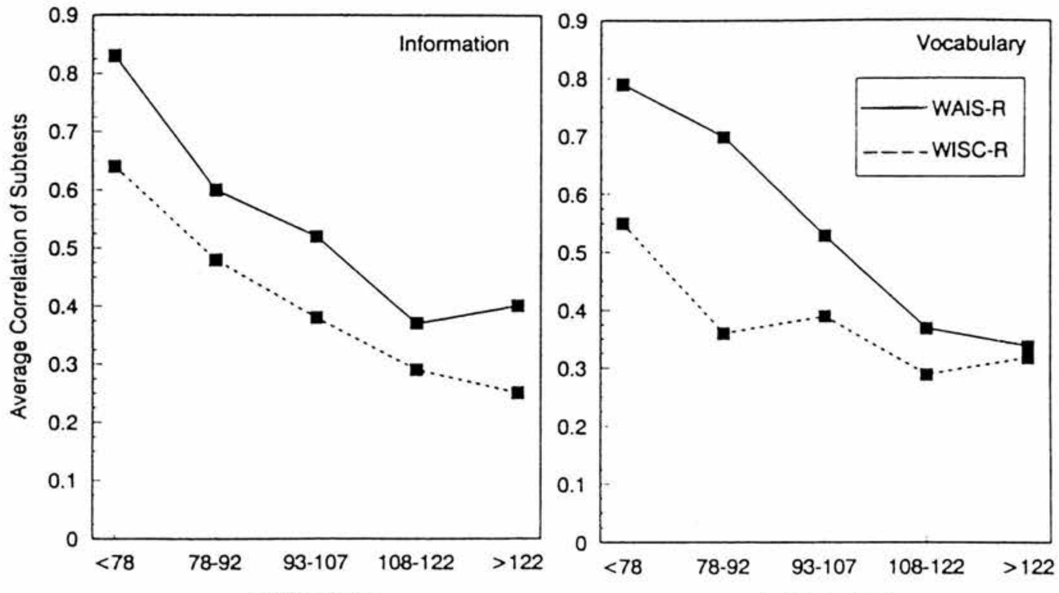
"Clearly, the correlations between cognitive task measures and WAIS-R IQ are up to twice as large in low IQ samples as in high IQ samples." (p.352)

Detterman and Daniel's next study was based on standardisation data from the WAIS-R and WISC-R. The full standardisation samples for the these test batteries consisted of 1,880 and 2,200 subjects, respectively. Five subgroups, varying in mean IQ, were selected from each of the standardisation samples on the basis of standard score on one or other of two sub-tests (Vocabulary and Information). This method of subject selection was chosen in preference to using full-scale score, since the latter has the effect of reducing inter-test correlations to an unacceptably low level because higher scores on one subtest will tend to cancel-out lower scores on another subtest.

Correlation matrices were constructed for each of the five groups tested with each battery. Two sets of results were calculated in each case, since samples were selected using either of two sub-tests. Correlations were averaged after correction for range restriction. The results of these analyses are shown in figure 3.2.

Figure 3.2 (From Detterman and Daniel, 1989)

Average correlation among WAIS-R and WISC-R subtests within ability level group when groups are selected by Vocabulary or Information subtests corrected for restriction of range.



The authors write of these findings:

> "The most obvious and striking trend apparent in figure (3.2) is that low ability groups demonstrate correlations which are two times larger than high ability groups...It is also apparent that there is a systematic trend for successively lower ability levels to show higher correlations." (p.354)

Detterman and Daniel's methods of analysis will be discussed in greater detail in a later chapter.

Lynn (1992) replicated Detterman and Daniel's study, using a sample of 1,369 Scottish children tested with the WISC-R. Subjects were divided in to five IQ bands, similar to those in the previous study. Mean intersubtest correlations declined from low to high IQ as follows: .44, .38, .17, 14, .20. Subsequent reanalyses of standardisation data from France and Japan revealed similar patterns of results (Lynn and Cooper, 1993; Lynn and Cooper, 1994).

Fogarty and Stankov (1995) argue that the differentiation effect observed by Spearman and, latterly, Detterman and Daniel, is a result of test limitations. *"statistical artifacts have not been eliminated as plausible accounts of trends in correlations"* (p.159) They believe that previous studies which have indicated a weakening of g at high IQ levels used tests which do not discriminate well between subjects in the upper ranges. They suggest that abilities may either diverge or converge at higher IQ levels and that this depends on the types of tests which are used.
In one of study, Fogarty and Stankov compared inter-test correlations in samples with low (88) and high (112) IQ on the Differential Aptitude Test (N = 20 and 25, respectively). Subjects were asked to perform *competing tasks.* These require the simultaneous performance of more than one sub-

task and are heavily 'resource demanding' and therefore difficult for high IQ subjects. The amount of common variance among the tasks was found to be greater in the high ability group, contrary to the law of diminishing returns. The authors speculate that, like the opposite of Spearman's analogy, this is akin to the situation where a strong person who is trying to lift a very heavy weight will benefit from additional energy but a weak person will never lift it, no matter how much more energy they get.
In another study, groups with low (92-103) and high (125+) IQ (N = 28 and 29) were given four perceptual speed tests. Such tests have no ceiling and exhibit equal discriminatory power at all ranges of IQ. Correlations between these measures were very similar in size in the low and high IQ groups.

Legree, Piper and Grafton (1996) argue that Fogarty and Stankov's findings arose because they used *achievement* tests, rather than ability tests. They note that Detterman (1993) was also unable to replicate the differentiation effect when using achievement tests and they cite a personal communication in which Detterman maintains that *"achievement tests do not exhibit a consistent pattern of correlations when the standardisation sample is divided into performance levels; instead the mean correlations...increase and decrease unpredictably"* (Legree et al., 1996, p.46).
Legree et al. (1996) turn around the argument that differentiation is a result of test bias, by claiming that the effect is *more* likely to be found where psychometrically sound measures are used. They divided the normative sample (N = 9,173) for the 1980 Armed Services Vocational Aptitude Battery into five groups, differing in ability level, but approximately equal in range. Divisions were made on the basis of each of the ten tests in the battery. This yielded ten sets of five correlation

matrices. Scores on seven of the ten scales were normally distributed but three showed severe skewness due to ceiling effects. When the tests with good psychometric properties had been used to define the samples the "decreasing positive manifold effect" was observed. In contrast, the pattern of mean correlations in samples selected by the psychometrically problematic tests showed no clear trend. The authors claim that this is because the latter tests *"lack the power required to identify groups of participants with similar levels of measurement error"* resulting in *"some groups being more heterogeneous than is indicated by the observed variance estimates"*. (p.53)

The most recent study of differention to have been published is that of Deary et al. (1996), who reanalysed the normative data from the Irish standardisation of the Differential Aptitude Test. The full sample comprised over 10,500 schoolchildren aged between 14 and 17 years of age. The study aimed to test both differentiation hypotheses, i.e. those relating to decreasing positive manifold with increasing age (in childhood) and increasing intelligence level. To do this they selected four groups, from the larger sample, having the following characteristics:
younger + lower IQ; younger + higher IQ; older + lower IQ, older + higher IQ
A special statistical procedure was created to select 'ideal' groups with widely separated means of IQ (mean 90 versus 110) and age (mean 170 versus 201 months) but near identical distributions. In each case, ability selection was made on the basis of score on each of the eight sub-tests separately. Matrices of intercorrelations amongst the remaining tests were computed in each of these cases. The selection process had the effect of reducing sample sizes such that the number of subjects in each group did not exceed the low hundreds. Correlations were averaged for each of the four groups, selected by each of the sub-tests, producing eight

matrices of correlations for each group. Each of these was subjected to principal components analysis. When the variance accounted for by the first principal component in each of the matrices produced for each group was averaged, the following pattern emerged:

Table 3.5 Average Variance Explained by the First Principal Component Across Eight Matrices in Four Samples Differing in Mean Age and IQ (From Deary et al., 1996)

Young/Low	Young/High	Old/Low	Old/High
49.8	47.8	49.9	47.8

These figures imply support for the hypothesis of reduced positive manifold at higher IQs but not the hypothesis of a stronger g factor in younger, as compared to older children. The authors concur with this interpretation. Close inspection of the individual results, however, reveals a less clear picture. It is only when samples are selected by three of the eight tests that g is stronger in the group with lower IQ. In the other cases the difference is negligible or in the opposite direction. Deary et al.'s methods and results will be discussed further in a later chapter.

Converging evidence in support of the differentiation hypothesis comes from studies of creativity. This will now be examined.

Creativity research supporting the differentiation hypothesis.

The subject of creativity has been a source of some controversy within differential psychology since around 1950. Some (e.g. Burt, 1955,1962) have argued that, with the exception of genius or in particularly specialised fields, individual differences in creative achievement can be accounted for by g. Others (e.g. MacKinnon, 1962; Torrance, 1962) maintain that intelligence tests are poor predictors of creativity, therefore the traits must be separate. This controversy can be closely associated with the dispute over the structure of intelligence and may be similarly resolved.

One of the most widely cited studies of creativity is that of Getzels and Jackson (1962). Theirs was an attempt to *"examine empirically the consequences of applying other conceptions of giftedness as well as 'high IQ' to the study of children"*. 535 school children were tested using five measures of creativity, some of which were adapted from tests developed by Guilford and by Cattell, and others specially constructed by the authors. (Word association, uses for things, hidden shapes, fables, and make-up problems.) Correlations between these measures and IQ were calculated for 292 boys and 241 girls separately. All correlations were positive and of moderate size, between +.10 and +.50, those for the girls being slightly higher. Correlations between the five creativity tests were generally higher than those between creativity and IQ, suggesting (as Guilford had hypothesised) the possibility of separable factors (although no factor analysis was performed). Next, the sample was divided into two contrasting groups, one composed of children having scored very highly on measures of intelligence, and relatively low on tests of creativity, and the other of those who had scored high on tests of creativity and relatively low on the tests of intelligence. Subjects scoring in the top

20% on the composite creativity measure were then compared to those scoring in the top 20% on the IQ measure. This selection had the effect of reducing the samples in number to 26 in the 'high creativity' group and 28 in the 'high IQ' group. The most interesting result was that the high creativity group equalled the high IQ group in scholastic achievement in spite of having an average IQ 23 points lower.
Getzels and Jackson interpreted their findings as evidence that creativity and intelligence are separable. Their sample, however, was profoundly atypical. Subjects were drawn from an exclusive private school in which a large proportion of the pupils came from the families of lecturers at the University of Chicago, and only a negligible proportion from the families of semi-skilled or unskilled workers. Their mean IQ was 132, and the mean IQs of the 'low IQ' and 'high IQ groups were 127 and 150, respectively. A replication of the study, using a more typical sample, yielded rather different results. Hasan and Butcher (1966) closely replicated Getzels and Jackson's study using 175 Scottish children having an equal range of IQ, but a mean 25 points lower. Measures of intelligence and creativity correlated to a far greater extent in the Scottish sample. For example, while Getzels and Jackson had reported a correlation of +.31 between score on the 'Fables' test and IQ (boys only), the corresponding correlation found by Hasan and Butcher was +.73. Not all the discrepancies were as large as this, but they were all in the same direction and all substantial. Unlike the Chicago result, IQ correlated more highly with total creativity score than did 9 out of 10 of the separate creativity tests used, indicating a far stronger general factor.

The suggestion by Burt (1955,1962) and others, that tests of creativity are essentially tests of g in the normal IQ range, would appear to be supported by Hasan and Butcher's results. The lower correlations found by

Getzels and Jackson would suggest that abilities associated with 'creativity' differentiate from g at high levels of intelligence, in line with the differentiation hypothesis. Likewise, Getzels and Jackson's finding of equal scholastic achievement in the 'high creativity' and 'high IQ' groups suggest that at high ability levels IQ tests have less predictive value in terms of school performance. (It should be noted, however, that a large number of subjects had to be excluded by Getzel and Jackson in order to form their high IQ/low creativity and low IQ/high creativity groups. This suggests that the influence of g in the whole sample was still clear, albeit weaker than in Hasan and Butcher's sample.)

Other studies of creativity have yielded similar patterns of results. For example, Mackinnon (1962) studied a number of high achievers in the fields of creative writing, architecture, mathematics, industrial research, physical science, and engineering. The mean IQ of these subjects was 113, using Terman's Concept Mastery Scale. MacKinnon writes of his findings:

> "As for the relation between intelligence and creativity, save for the mathematicians ,where there is a low and positive correlation between intelligence and level of creativeness, we have found within our samples essentially zero relationship between the two variables (-.08)." (p.491)

Cropely (1966) studied a more representative sample and found a significant correlation between creativity and intelligence.

A study by Hargreaves and Bolton (1972) yielded particularly interesting results, from the point of view of the differentiation hypothesis. These investigators divided their total sample (mean IQ 102) into three range-matched groups of high, low and intermediate ability. Correlations between measures of intelligence and creativity fell from .46 in the low

IQ group to .09 in the high IQ group (mean IQ 122). In the high IQ group correlations between IQ and total creativity were lower than those between total creativity and the separate tests of divergent thinking, whilst the reverse was true in the low and intermediate groups.

The explanation for these findings which has the greatest significance for the differentiation hypothesis, is the *threshold theory*, labelled this way by Mackinnon (1962) but described by many previous and subsequent investigators of creativity (eg. McClelland,1958; Barron,1963; Hasan and Butcher,1966; Meer and Stein,1955; Yamamoto, 1964; Hudson,1965; Moore,1966; Hargreaves & Bolton,1972). Briefly, the threshold theory suggests that above a certain required minimum level of intelligence, which varies from field to field and in some cases may be surprisingly low, being more intelligent does not guarantee a corresponding increase in creativeness. Barron (1963) has suggested a threshold of IQ 120, above which intelligence ceases to be so relevant to many forms of achievement. If this suggestion is correct then it would account for the fact that in most individuals creativity measures are, like measures of specific cognitive abilities, highly g-loaded, whereas in high-IQ samples, scores on such measures appear to exhibit a greater degree of independence from the g factor. Since the majority of individuals within the population are below this criterion, studies of representative samples would be expected to reveal a strong g factor, as indeed they have. [Parallels between the threshold theory and the theories put forward to explain factor differentiation at high IQs will, no doubt, be clear to the reader.]

Evidence from research involving measures of mental speed.

Studies involving psychophysical correlates of intelligence also provide some incidental evidence for the hypothesis. A number of researchers working with Inspection Time (IT), for example, have reported higher correlations with IQ in samples of lower, as compared to higher, ability.

For example, in their original study, Nettelbeck and Lally (1976) found a correlation of -.92 between IT and Performance IQ on the Weschler Adult Intelligence Scale. This very high correlation has been difficult to replicate and has been attributed to the small sample size (N=10) and very large ability range (borderline retarded to high IQ) (eg.Mackintosh,1981). Although other researchers have reported stong correlations (e.g. Lally and Nettelbeck, 1977; Brand and Deary,1982) these have tended to be markedly lower in samples which have excluded retarded subjects. Many studies have, in fact, obtained low or non-significant correlations (e.g Hulme and Turnbull,1983; Irwin,1984; Nettlebeck, 1982; Smith and Stanley, 1983; Vernon,1983). When Nettelbeck and Kirby (1983) correlated IQ and IT in a sample of 91 subjects with a more normal distribution of IQ the correlation between IQ and IT was markedly lower than has been the case in their original sample (-.50).
Although selection effects may well explain larger correlations in groups containing retarded subjects, it may be speculated that differentiation effects also contribute.

CHAPTER 4

GENERAL METHODOLOGY FOR THE PSYCHOMETRIC STUDIES

This chapter will describe the general methodology used in the psychometric investigations of the age and IQ differentiation hypotheses. It was thought appropriate to supply this information in one chapter, since the same psychometric tests were employed in both investigations and since the subjects of these studies are drawn from the same overall sample. The need for undue repetition in later experimental chapters is thus circumvented. This is particularly important given the lengthy nature of the information.

The main aims of this chapter are as follows:

to describe the rationale underlying the selection of the sample

to describe the sample

to explain the reasoning behind the selection of the psychometric measures in the test battery

to describe the tests in the battery and to provide procedural details regarding their administration.

to evaluate the degree to which the sample is representative of the population for IQ.

THE SAMPLE

Key Sampling Considerations

The identification of appropriate subjects is perhaps the most important consideration in any investigation of the differentiation hypotheses. The two main considerations which must be taken into account are the age and ability ranges of the subjects, both of which should be as wide as possible. In particular, the sample should be broadly representative of the general population for ability. This is important because a) this is more likely to produce results which can be generalised to other samples, b) such sampling is necessary if variable distributions are to meet good psychometric criteria and c) the distribution should be wide enough to allow for meaningful sub-samples to be drawn from disparate parts of the ability distribution for the purposes of comparison.

Preparatory Work

In an effort to meet the latter criterion, the initial plan was to test subjects from non-selective state schools whom it was thought would represent the widest possible ability range.

Before approaching schools directly, it is customary to gain permission from the local education authority. Detailed proposals were submitted to the LEA, however the number of channels and committees through which these had to pass was far more elaborate than had been anticipated and several months elapsed before the initial stages were completed. For this reason the decision was made to contact an independent-sector school which had cooperated with the psychology department in the past.

Features of the Chosen School

Although private, the chosen school is largely non-selective at the primary and early secondary levels*, and the general ethos is towards mixed ability teaching. The school is also well regarded for its efforts in remedial teaching and is, for this reason, actively chosen by parents of children with learning disabilities (including low IQ). The strong academic reputation of the school is also a criterion for parents of high ability children. It was thus anticipated that the ability range of the sample would be wide enough to be considered representative of the general population, although it was expected that mean IQ would be slightly higher than average.

Average class sizes at the school are smaller than those in the state sector, varying between 23 and 28. Classes are co-educational at the primary levels and single-sex at the secondary level, where girls and boys move to separate campuses.

[*Some selection occurs for new pupils at the secondary level, however the majority of secondary pupils are continuing students from the non-selected primary classes. Incidental selection pressures will, of course, operate in such a school, since the fees are not insubstantial and parents' occupations, and IQs are likely to be in the upper bracket. Nevertheless, it was considered an acceptable compromise, given the lack of available alternatives. There are, of course, many other advantages of using a private school, not least good student discipline and provision of testing rooms for the individual testing described in chapter 8.]

Ages and School Grades Represented in the Sample

The actual age range selected for study was decided on the basis of three main considerations:

1) Several school grades should be covered and these should relate to the years during which ability levels tend to increase appreciably.

The primary and early secondary grades were considered suitable.

2) It is essential that the same tests can be used with all the subjects.

Thus age range selection was constrained by the availability of suitable test instruments.

3) The requirements of the schools must be met.

The key restriction of the school was that grades involved in preparation for important examinations would not be available for testing. *

In the light of the above, the age range *eight to twelve years* was considered appropriate. This range is represented in *Primary grades 4 to 7 and Secondary 1.*

*[The school contacted pupils' parents to ask for their consent to the testing. Form and class teachers were also asked for their agreement. To the best of my knowledge, no parents or teachers objected.]

Sample Sizes

In all, 549 children were tested with the chosen psychometric battery. Four classes of pupils were tested at each age except age nine, where class sizes were smaller than average, requiring five classes to be tested for the within-age N to exceed 100.

Numbers of subjects in each school grade are shown in Table 1. In the main, students at each grade are the same chronological age. The reader will note, however, that this is not always the case. A common policy of the school is for low achievers to be kept behind a grade and for very high achievers to be accellerated. As a result each grade will contain a number of younger, brighter-than average children and older, low achievers.

One of the conditions stipulated by the school was that individual pupils would be be excused from the testing sessions to attend pre-scheduled music lessons. This resulted in a number of subjects with missing scores on one or more tests. Since missing scores render a subject's other marks unusable in correlational analyses, these subjects were excluded from further consideration. This left 538 full sets of scores.

Table 4.1 Sample Sizes at Each School Grade

GRADE	Total sample size	AGE RANGE (months)	N excluding subjects with missing scores	Number of Ss at age for grade
P4.	113	87-114	112	[AGE 8 =104 -> 103]
P5.	130	99-128	125	[AGE 9 = 110 ->106]
P6.	100	109-138	99	[AGE 10 = 85 -> 84]
P7.	105	129-143	102	[AGE 11 = 96 ->93]
S1.	101	141-161	100	[AGE 12 = 93]
	Total N = 549		Total number of Ss with full sets of results = 538	[N Ss where age & grade match = 479]

Table 4.2 records the numbers of children from the sample at each *chronological age*. [As mentioned above, CA and grade are not always consistent.] A number of subjects at the lowest and highest grades were younger than 8 or older than 12. Scores for these subjects were excluded from further analyses along with those for subjects with missing values. The total number of subjects remaining after these exclusions is 525.

Table 4.2 Sample Sizes at Each Chronological Age .

AGE	Total sample size.	N excluding subjects with missing scores.
EIGHT	113	111
NINE	118	114
TEN	99	98
ELEVEN	108	104
TWELVE	98	98
	[N tot= 536]	*[N tot= 525]
		*=Final numbers included in analyses

(One 11 yr old/ Grade 7 boy was later excluded since his RPM score was below the lowest percentile on the norm tables. This reduced the total sample N to 524)

SELECTION OF PSYCHOMETRIC TESTS

Burt (1954) and Anastasi (1964) made a number of suggestions concerning the design of studies aimed at testing the differentiation hypothesis.* These were followed when selecting psychometric measures.

Breadth of variable sampling

Burt and Anastasi recommend that the tests chosen should be such as to elicit group factors unambiguously. In other words, the test battery should be such that differentiation will show up if it is there. A battery comprising only tests of the same ability (e.g. information or verbal comprehension) will tend to be dominated by one common factor, with little variance remaining, after it has been extracted, for meaningful comparisons to be made between factor structures at different ages or IQ levels.

Burt and Anastasi also recommend that there should be more than one test representing each factor to be identified. This criterion may be difficult to meet in the sense that it is not always obvious, prior to data analysis, what the factors will be. Nevertheless, the psychometric literature is a good guide to the types of tests which tend to be common to particular factors.

The chosen tests will now be described.

* These recommendations were primarily directed towards studies involving comparisons between age groups, however, many are equally appropriate for studies of differences across the IQ spectrum.

TESTS OF GROUP AND PRIMARY FACTOR LEVEL ABILITIES

Background to Thurstone's Tests of the Primary Mental Abilities

Since L.L.Thurstone was the original proponent of the multifactor approach to intelligence it seems appropriate that the tests developed by him should be used in an investigation of ability differentiation. The tests of the Primary Mental Abilities (originally the Chicago tests of the PMAs; Thurstone,1941) were derived directly from Thurstone's original factor analysis of abilities. The tests were designed to measure, as closely as possible, Thurstone's five hypothetically independent primary factors. In this respect they represent the ideal measures referred to by Burt (1954) in respect of the differentiation hypothesis. The tests were subsequently refined and restandardized in 1962 (T.G. Thurstone, 1963). By this time, however, the importance of a "second-order general factor" common to all the tests had been clearly acknowleged. Thus, in the manual for the 1963 edition, global test score is recommended as "a reliable estimate of intelligence comparable to scores on tests such as the Stanford-Binet and the Weschsler Intelligence Scale for Children" (Thurstone, 1963, p.4).

Description of The Primary Mental Abilities Tests

The PMA test battery consists a set of multiple choice subtests designed to be used in a group-testing situation. Eight subtests contribute to the 5 PMAs. Each of these yields an IQ and these can be averaged to yield a global IQ score.

VERBAL MEANING

Handbook definition: "The ability to understand ideas expressed in words".

The Verbal Meaning test comprises two, separately administered, 30 item subtests which differ in difficulty level:
The easier of the two subtests is a picture vocabulary test. The testee must select, from an array of 4 pictures, the one *named* by the tester. [e.g. experimenter: "Question 1 - Find the DOG"]
The second subtest is a verbal comprehension task. For each item the child must select, from 4 alternatives, the word which 'means the same as' the *printed* target word.

Given the differing difficulty levels of the two subtests, the Verbal Meaning test score must be considered whole for it to provide meaningful information across ages and ability levels.

NUMBER FACILITY

Handbook definition " The ability to work with numbers, to handle simple quantitative problems rapidly and accurately, and to understand and recognize quantitative differences."
The Number Facility test also comprises two subtests. The first of these (Number Sense) comprises 10 series completion items and 10 realistic numerical reasoning problems.
The second subtest (Addition) consists of 30, three-number addition tasks. Since there are more addition items than can be completed by most testees in the time given, this assesses both accuracy and speed.

As with the Verbal Meaning test, the subtests differ in difficulty; Number Sense being the easier of the two. Thus, for age comparisons, only scores for the *whole* test provide meaningful information without ceiling effects.

SPATIAL RELATIONS

Handbook definition: "The ability to visualize objects and figures rotated in space and the relations between them."

The spatial relations test involves 25 mental manipulation problems. A target item is presented as a 'square with a piece missing' and the task is to choose the missing piece from an array of four possibilities.

REASONING

Handbook definition: "The ability to solve logical problems".
The reasoning test also comprises two subtests. Unlike the subtests of the VM & NF scales, however, these are of equivalent difficulty level and are appreciably different in terms of content. For this reason the test scores can be used meaningfully as independent measures, and they have been separated in many of the analyses which follow. In brief, the scales are as follows:

Figure Grouping: The task is to choose, from a series of abstract figures, the one which is 'the odd one out'. Whilst this is primarily a test of abstract reasoning ability it involves visuospatial skills to a large degree.

Word Grouping: Again, the task is to find the odd one out, but the stimuli are words. In addition to logical reasoning, this test taps verbal comprehension skills.

Each of the reasoning sub-tests contains 25 items.

PERCEPTUAL SPEED

Handbook definition " The ability to recognize likenesses and differences between objects or symbols quickly and accurately."

Each item on the Perceptual Speed test involves an array of four abstract figures, two of which are identical. Subjects must complete a simple visual scan and comparison to select the two identical figures. Subjects are not expected to be able to complete all 40 items in the time given.

Tests of Specific Abilities

'Creativity' tests:

Given their relevance to the differentiation hypothesis (see chapter 3) the decision was made to include two tests of divergent thinking in the test battery.

WORD FLUENCY

The first of these is the Word Fluency test - a test of ideational fluency. This first appeared as a test of verbal ability in the earliest edition of Thurstone's Primary Mental Abilities Test, indeed it was identified as one of the "seven primary mental abilities of intelligence" (Thurstone and Thurstone, 1941). Later work - which reduced the number of PMAs to 5 plus a second-order general factor (as in the 1962 version used in the present study) - revealed Word Fluency to be a less pure measure of verbal ability than had been thought. In fact, it has more in common with tests of divergent production - central to what is often called 'creativity' (Guilford 1967; Guilford & Hoepfner, 1971).

Word fluency tasks can take many forms - One example is the 'word beginnings and ending test' where the task is to write down as many words as can be thought of which start and end with a certain letter or combination of letters, or with a particular prefix or suffix.
The current study employed a simple category production task -"Write down the names of as many animals as you can think of ". Total score is simply the number of different names produced.

FIGURES OF SPEECH

The second 'creativity' test - the Figures of Speech test - was drawn from the Princeton *Kit of Factor Referenced Cognitive Tests* (Ekstrom, French and Harman, 1976). This test also also addresses divergent production, or fluency; but rather than category production it involves the generation of associated ideas.

Items involve open-ended descriptions associated with figures of speech such as "Her hair was as red as..." "The jewels sparkled like..." The task is to think of as many ending as possible (up to a maximum of three) for each item. Answers may be in single words or in phrases. Due to the difficulty of making qualitative judgements, items were scored simply on the basis of *number* of appropriate endings. There are 10 items in total (administered in two sets of 5) each of which each asks for up to three answers. Thus this is essentially a 30 item test.

This type of ability has been referred to as ideational fluency; expressional fluency, associational fluency or even DMU - the *D*ivergent Production of *SeM*antic *U*nits (e.g. Carrol, 1941; Guilford and Hoepfner,1971). The exact terminology for this type of test can be quibbled over, and some psychometricians - most notably Carroll (e.g.1993) have disagreed as to which 'factor' it belongs to. Nevertheless, the Figures of Speech test has common elements in all of the above and in most analyses (unless we are factoring a multitude of divergent production tests) these distinctions are redundant.

FORWARD DIGIT SPAN

Digit span is another specific-factor level variable which bears on the differentiation hypothesis. Although often referred to as a *cognitive correlate* of psychometric intelligence, this short-term memory test regularly appears in psychometric test batteries as a supplement to the more usual pen-and-paper ability tests. For example, it appears in the WISC-R (Wechsler,1974) and the Stanford Binet (Terman & Merril, 1960). The digit strings used in the present study were adapted from the lists supplied by Ekstrom, French and Harman (1974) in their *Kit of Factor-Referenced Cognitive Tests* (Auditory number span test; MS1), although there is nothing particularly unique about the procedure.

Lists of numbers varying in size were read out at a rate of one digit per second. Immediately following presentation of the final digit, subjects were given the cue 'BEGIN' which indicated that they should start writing down the list. For each set size, three lists were presented. Sets began at three digits and continued to nine digits. Span for a particular set size was said to have been achieved when at least two out of three sets were recalled accurately, in the correct order.

Since most of the aforementioned tests were designed to assess hypothetically independent factors, they meet Burt & Anastasi's criterion well. These recommendations were, however, specifically tailored towards Burt's age differentiation hypothesis. While they provide useful general guidelines they do not cover all the relevant considerations; particularly when the IQ differentiation/divergence hypothesis is considered. The essential premise of the differentiation hypothesis is that g (or whatever is responsible for g) is weaker in older/smarter subjects. Put another way, the predictive validity of individual tests will reduce as the age or IQ range is ascended. Since investigation of the hypotheses will involve selecting or drawing conclusions about subjects on the basis of their level of general intelligence, we must be confident that our selector is a valid measure of general intelligence - even in the most hypothetically differentiated groups. If the battery contains only tests of primary and specific factors, we cannot be entirely confident about the validity of our supposed ability splits, particularly in the oldest/brightest groups. Thus, the present investgations demand an independent measure which is highly g-loaded at all ages and ability levels. [Another reason for including a test of general intelligence is that it acts as a benchmark by which to evaluate other variables in a factor analysis.]

A Unitary test of General Intelligence: Raven's Matrices

The unitary test of general intelligence which is most widely cited by researchers in the field of intelligence is the Raven's Standard Progressive Matrices (RPM/SPM) test.

Raven's matrices (coloured, standard, and advanced) are measures of nonverbal abstract reasoning ability, closely identified with fluid intelligence (Kline, 1993; Carrol, 1993). The Standard Progressive Matrices (SPM) is the most widely used. Each item consists of a matrix of patterns or abstract figures which follow one or more rules. Rules can be deduced by reading across the rows of the matrix or down the columns. One of the six figures from the matrix is missing, and the task is to select, from a set of eight options, the piece which completes the sequence. The 60 items increase in difficulty, beginning with the fairly concrete and culminating in the highly abstract. Early items involve simple pattern recognition and visuospatial comparison. Later items involve rule comprehension and logical reasoning. Because the RPM is a nonverbal test and its early items involve pattern recognition, some researchers have referred to it as a test of spatial ability. This is not strictly true, since RPM scores are correlated highly with verbal as well as nonverbal items. Nevertheless, early items do load on visuospatial/ perceptual factors and although variance in later items is largely explained by g, it is likely that visuospatial group factors continue to play a role.

Discriminatory Power

Burt and Anastasi also recommend that the tests chosen should not have restrictive upper or lower limits since this may result in artificial restriction of range.

Attention must be paid to the groups upon which the tests were standardized. A suitable test will have good discriminatory power over a wider range of abilities and ages than those expected in the sample. Tests for which there is no effective 'last item' are also less prone to ceiling effects, particularly when the time allowed exceeds the likely number of answers given.

Thurstone's PMA tests.

The PMA tests [IBM 805 Edition for grades 4-6] were standardized in 1962 on a sample of over 6000 children aged between 8 years, 4 months, and 13 years, 9 months. The IQ ranges appearing in the norm tables vary between IQ 50 (lowest score, oldest sample) and IQ 181 (highest score, youngest sample)*. Given this purported discriminatory power over such a wide range of mental ages it was anticipated that the PMA tests would be suitably free of ceiling effects in a sample aged 8 to 12 , even if (as suggested by Flynn, 1987) there has been a rise in intelligence since 1962.

Raven's SPM.

The Raven's Matrices Test is also suitable for a wide range of abilities and ages. The manual suggests that the test can be used with subjects aged from 6 to 65 years. The 1979 British normative data extends to the 100th percentile for a population of subjects aged up to 15 years, 8 months. Again, no ceiling was anticipated.

* [Deviation IQ with mean = 100 and SD = 15].

Word Fluency & Figures of Speech

The Word Fluency test is an open-ended test not constrained by a fixed upper limit. Previous research by the current author (e.g. Lynn, Chan and Pagliari, 1988) suggests that most subjects reach their maximum before the time limit expires.

As with the verbal fluency test, score on the Figures of Speech test is a function of number of ideas generated, rather than number of items. It was not anticipated that subjects would be able to think of as many possible responses as there were response opportunities (i.e. 30) and it was expected that the time limit would exceed that required for maximum performance.

Digit Span

Digit span sets were continued beyond the capabilities of all subjects, thus the task had no ceiling.

Test Reliability and Validity

When selecting tests it is also important to consider their reliability and validity. Although test reliability may be affected by factors such as age (Anastasi, 1988) good baseline reliability is a useful guide.

The PMA tests

The PMA tests have been used in numerous published studies since the 1940s (e.g. Lynn, 1990; Schaie & Herzog, 1983) and Kline (1993) reports that they are reasonably valid measures of the primary mental abilities.

Unfortunately, no test reliability details are supplied in the Manual for the 1963 edition of the PMA tests. According to Kline (1993) and Carroll (1993) few, if any, studies have directly addressed this issue. Data from the 1941 edition may, however, be used as a guide. The following split-half reliabilities were recorded:

Table 4.3 Split half reliability coefficients for four scales of Thurstone's Primary Mental Abilities Test - 1941 Edition.

Verbal	.94
Reasoning	.92
Space	.90
Number	.92

It is plausible to assume that later editions were even more rigorously designed.

Raven's Matrices

Validity and reliability data for Raven's Standard Progressive Matrices are much more widely available than that for the PMA tests. According to most psychometricians, this test has very high validity and reliability, with internal consistency estimates typically exceeding r = .90 (Kline, 1993)

Other Tests

Estimates of internal consistency reliability are not of great meaning when considering tests such as Word Fluency, Figures of Speech and Digit Span. In practice these measures are likely to be highly consistent since each is a pure measure of a very specific skill.

[The issue of test reliability and validity will be returned to in later chapters]

TESTING PROCEDURES

Table 4.4 summarises the administration times for each test and the order of presentation.

Each test in the battery is accompanied by one or more practice items. These are particularly important for the youngest children, for whom test requirements may not be immediately clear. Clarity was improved by magnifying the practice items and answer sheets to poster-size.* This does not change the content of the message but simply makes the instructions clearer for younger subjects and those sitting at the back of the testing room. Slight modifications in the instructions for two tests allowed for demonstration rather than simply description. For example, the magnified practice item on the Raven's Matrices test was modified such that the 'missing piece' could be lifted from the array of potential solutions and placed in the blank space on the main matrix. Likewise, in the Spatial Relations test the four potential missing pieces of the partial target square could be lifted like jigsaw pieces, to demonstrate how they did or did not complete the square. These illustrations were, of course, of most benefit for the youngest subjects - instructions are self-evident to older testees.

The times listed as being required for directions are those given in the handbooks. All subjects received the verbatim instructions suggested in the test manuals however these were delivered more slowly to the youngest subjects and were elaborated with the posters mentioned above. For this and other reasons (see footnotes overleaf) total testing time was around half an hour less in the oldest subjects.

* Since the PMA tests are out of print, original score sheets were unavailable. New scoring forms were prepared by the experimenter and reproduced by the technical services department of Edinburgh University Psychology Department. Answers are selected by placing a cross in one of four boxes labelled a) to d).

TABLE 4.4 TESTING SEQUENCE AND ADMINISTRATION TIMES.

TEST	TIME LIMIT (Minutes)	DIRECTIONS etc. (mins. approx.)
Raven's Matrices[1]	45	10
PMA Verbal Meaning		
Words	7	5
Pictures	6	5
PMA Number Facility		
Number Sense	10	5
Addition	4	5
PMA Spatial Relations	6	5
PMA Reasoning		
Figure Grouping	8	5
Word Grouping	6	5
PMA Perceptual Speed	5	5
Verbal Fluency	2	2
Digit Span	5	2
Figures of Speech	8	4
TOTAL TIME	112	58
REST PERIODS etc.	5-15 mins[2]	
TOTAL TESTING PERIOD:	c. 2.5 - 3 HOURS	

[1] Although 45 minutes are allowed for Raven's Standard Progressive Matrices this tends to be an overestimate of the time needed to complete the items. Subjects are instructed to take as long as they wish, with the stipulation that they complete all the items sequentially. Older grades completed the test in a shorter time period than allowed. Once all testees were satisfied with their answers and had closed their question books, the session was terminated and the new test begun.

[2] The number and duration of rest periods was greater for the younger subjects.

EVALUATING THE SAMPLE

Before considering the evidence for differentiation it is useful to evaluate the intellectual standing of the sample, relative to the general population. To this end, standardized IQs were calculated for the Raven's Matrices and Primary Mental Abilities tests using published norm tables. (Normative data are from the 1979 standardization of the SPM and the 1962 standardization of the PMA tests.) Suitable standardization data for the other tests are not available. Table 4.5 shows the mean Raven's Matrices IQ for each age group and for the sample as a whole.

Table 4.5 Mean and Standard Deviation of Raven's Matrices IQ, Relative to 1979 British Norms.

AGE	SAMPLE SIZE	DEVIATION IQ	S.D.
8	[N=111]	109.74	15.40
9	[N=114]	111.28	15.90
10	[N=98]	112.96	15.00
11	[N=103]	115.33	14.30
12	[N = 98]	115.28	12.70
ALL	524	112.80	14.90

Tables 4.6 a) to e), overleaf, show the sample IQs for Thurstone's tests of the five Primary Mental Abilities, at each age. Table 4.7 presents a summary of the PMA IQ data for the total sample.

[It will be noted that the sample of 11 year olds has been reduced to N=103, following removal of one boy whose RPM score was below the lowest percentile on the norm tables and thus could not be converted to an IQ. This subject's other test scores were also lower than average, and the mean of his PMA IQs (84.2) was the lowest of all subjects, irrespective of age. The RPM score (12) was particularly poor, however, suggesting that more than IQ was involved in the errors.]

Table 4.6 (a) PMA - VERBAL MEANING IQ

Age	Sample Size	IQ	S.D.
8	[N=111]	116.32	12.93
9	[N = 114]	118.41	14.15
10	[N = 98]	118.92	11.72
11	[N = 103]	119.67	11.70
12	[N = 98]	119.89	11.07
ALL		118.60	12.30

Table 4.6(b) PMA - NUMBER FACILITY IQ

Age	Sample Size	IQ	S.D.
8	[N=111]	108.06	11.26
9	[N=114]	110.32	9.05
10	[N = 98]	108.49	8.19
11	[N = 103]	104.85	8.23
12	[N = 98]	103.82	8.88
ALL		107.20	9.50

Table 4.6(c) PMA SPATIAL RELATIONS IQ

Age	Sample Size	IQ	S.D.
8	[N=111]	107.297	15.165
9	[N=114]	108.877	13.941
10	[N = 98]	114.776	13.21
11	[N = 103]	115.340	16.32
12	[N = 98]	117.286	14.434
ALL		112.50	15.10

Table 4.6 (d) PMA REASONING IQ

Age	Sample Size	IQ	S.D.
8	[N=111]	107.54	13.31
9	[N=114]	112.23	13.97
10	[N = 98]	113.41	14.91
11	[N = 103]	114.04	12.17
12	[N = 98]	116.93	11.43
ALL		112.70	13.50

Table 4.6(e) PMA PERCEPTUAL SPEED IQ

Age	Sample Size	IQ	S.D.
8	[N= 111]	112.23	11.99
9	[N= 114]	112.14	10.41
10	[N = 98]	109.24	9.45
11	[N = 103]	110.71	9.45
12	[N = 98]	109.09	8.12
ALL		110.70	10.20

Table 4.7

AVERAGE PMA IQ AT EACH AGE

AGE	IQ
8	110.3
9	112.4
10	112.9
11	112.9
12	113.4
ALL	112.3

Evaluating the data in terms of current population norms.

According to Flynn (1987) general intelligence has increased by approximately 3 points per decade over the last half century. By this estimate, IQ should have increased by around 3.6 points between the standardization year of 1979 and 1991. This reduces the SPM IQ of the current sample to approximately 109, relative to the general population.

It is more difficult to estimate the relative standing of the sample with respect to individual PMA IQs. Informative data on this issue is, however, provided by Lynn (1990). Lynn reports findings for 307 fifth and sixth grade children (9-11 year olds) tested in 1978 and 310 children, from the same grades in the same school, tested in 1988. (Grades are combined.)

Lynn's data are reproduced, and compared to the current sample, in table

4.8 For the purposes of comparison, mean IQs for individual subtests and average PMA IQ have been calculated for a subsample of subjects from the same ages and grades as Lynn's, and for the total sample.

Table 4.8 Comparison of PMA IQs in the current sample with those reported by Lynn (1990).

	LYNN'S DATA 1978	LYNN'S DATA 1988	CURRENT SAMPLE (1991) Average for9-11 year olds in grades 5 & 6 [n = 213]	CURRENT SAMPLE (1991) Average of total sample [N =524]
VM	109.5 $\sigma = 18.2$	109.8 $\sigma = 18.2$	118.7 $\sigma = 13.0$	118.6 $\sigma = 12.3$
NF	104.6 $\sigma = 13.2$	107.2 $\sigma = 14.0$	109.3 $\sigma = 8.9$	107.2 $\sigma = 9.50$
SR	97.6 $\sigma = 16.8$	103.8 $\sigma = 16.1$	111.4 $\sigma = 13.9$	112.5 $\sigma = 15.1$
RSN	106.2 $\sigma = 15.5$	107.5 $\sigma = 16.2$	113.1 $\sigma = 14.5$	112.7 $\sigma = 13.5$
PS	98.6 $\sigma = 12.81$	107.5 $\sigma = 16.5$	110.8 $\sigma = 9.4$	110.7 $\sigma = 10.1$
Mean	103.3 15.3	107.1 16.2	112.7 11.9	112.3 12.1

Comparing total sample means to Lynn's most recent findings, we can estimate that the subjects in the current sample have the following approximate advantages over comparable state school pupils:

Verbal Comprehension:	8.8 Points
Number Facility:	0.0 Points
Spatial Relations:	8.7 Points
Reasoning:	5.2 Points
Perceptual Speed:	3.2 Points

The advantage in verbal ability is as expected, given the likely relative socioeconomic advantage of the children in the current sample. (c.f. Jensen & Reynolds, 1982). The advantage in spatial ability is more difficult to explain, however, particularly given the small advantages in other variables. (Lynn and others have suggested a nutritional theory to account for rises in nonverbal/fluid intelligence. Since SES relates to both quality of verbal interaction and and nutritional status it may explain both score advantages.)

Despite differences between sub-tests, if the PMA IQs are averaged to give an estimate of general intelligence, the resultant mean IQ is 112.3 This represents an *advantage of only around 5 points* relative to Lynn's obtained average of 107.1 (1988 data). The figure of 112.3 is also remarkably close to the mean Ravens Matrices IQ of 112.8, despite the difference in standardization year.

Lynn's finding of a 3.84 IQ point gain in general intelligence in the 10 year interval between 1978 and 1988, is in the order of that found by Flynn. If we assume that Lynn's population is representative for IQ, his estimate of 107.14 can be revised to 100. In the same way, assuming that Lynn's figures are correct, the mean of the current sample should be revised downwards by approximately 5 points (Lynn's estimate of annual gain, times 13 years) to 107.3. This compares to an IQ of 109.0 for the Raven's Matrices data revised according to Flynn's estimate.

The above calculations may appear unnecessarily complicated. Nevertheless, this process is essential if we wish to establish the true standing of the current sample, relative to the general population.

IQ Distribution

It is also important to consider the *distribution* of IQ in the current sample, in relation to that of the population. This can be estimated in one of three ways:

a) By comparing the standard deviation of the obtained RPM IQ with that expected in the population.

b) By averaging the SDs for the individual PMAs IQs and comparing this average to the expected population SD.

d) from some combination of a and b.

Since the population (represented by the standardization sample) is assumed to have an IQ standard deviation of 15, the obtained SD of 14.9 for the Raven's Matrices implies that the IQ distribution of the current sample is only fractionally less than that of the population.

This seems a conservative estimate, given the demographics of the sample, however, it appears less unlikely if we consider the 1979 standardization group. This "nationally representative sample of British school children" (N=3500) excluded children attending special schools, i.e. those with low IQs (as well as other problems). Thus the range of scores is not as wide would be expected if the full ability range had been sampled. For the reasons outlined in the last chapter, the current sample is likely to be widely distributed for IQ, despite the fact that the school is private.

By the second estimate, the average SD for PMA IQs[1] (=12.1), the score distribution of the current sample would appear to be approximately 19% narrower than that of the standardization sample[2]. The standardization group for the PMA was larger than that for the Standard Progressive Matrices (N=6370) and is likely to have included a wider range of abilities. The choice of which estimate to accept is essentially up to the reader. Averaging the two values, however, yields an overall estimate of 13.5 suggesting that the distribution is around 9% less than would be expected if the sample were fully representative of the population.

In summary, it would appear that the mean IQ of the current sample is approximately 8 points above that of the general population, or just over half of one standard deviation. The distribution for general intelligence, although broadly normal (see figs 4.1, 4.2 a, b), is around 10% narrower than

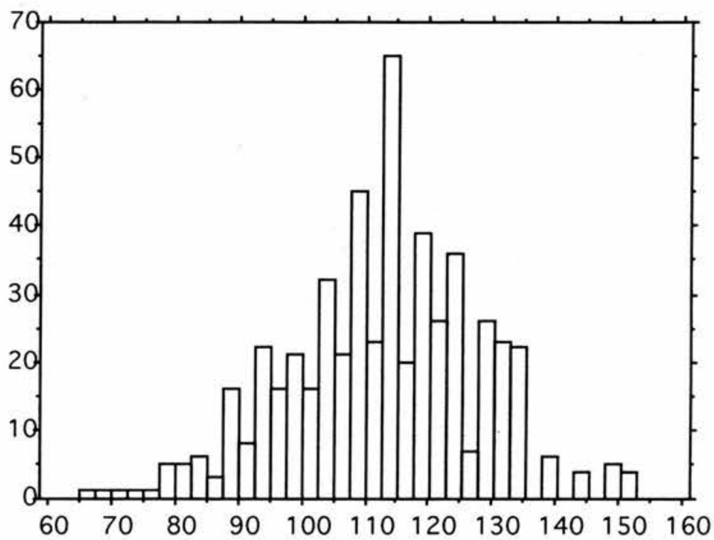
[1] Mean PMA IQ will be the same whether it has been derived by averaging the mean IQs for each of the 5 PMAs - as here- or by calculating each subject's mean PMA IQ and averaging these. The standard deviation will be less in the latter case, however, because it is a distribution of means, not real IQs. A better estimate of PMA distribution is provided by the method used - See figs 4.2 a & b.

[2] If Lynn's estimates are used for this comparison, the figures are as follows: 1978 average sd=15.3 (current sample 21% less), 1988 average sd= 16.2 (current sample 25% less).

would be expected if the sample were perfectly representative of the standardization group. Despite these differences from the ideal, the sample is considered to be suitable for the purposes of this project.

Figure 4.1 depicts the distribution of Raven's Matrices IQs in the full sample.

Figure 4.1. Distribution of Raven's Matrices IQ in the full sample

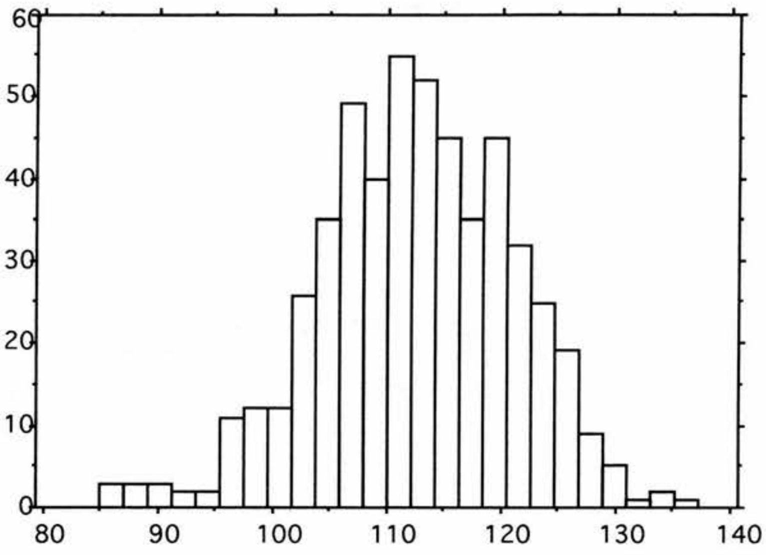


IQs exceeding 135 should be interpreted cautiously, due to difficulties in converting the highest raw scores to IQs using the available norm tables. Although the 1979 British norm tables are the most comprehensive versions available, they do not differentiate well within the 99th percentile, except for certain age groups. By the tables of areas under the normal curve, percentile 99 is equivalent to Z-scores 2.33 through 3.50, or IQ 135 to152 - a wide range. The next percentile listed is 100; approximately equivalent to IQ 152+ (Z scores over 3.50). To avoid a dramatic jump between IQs 135 and 152, a certain amount of estimation was involved. In most cases this simply involved using the estimates for the age group immediately above or below. This is unlikely to result in dramatic departures from the true IQ since norms are provided for every 5-month age interval.

Four subjects obtained scores which exceeded the reported 100th percentile for their age group. For the purposes of the present study, all of these have been allocated IQs of 152 (Z =3.50), although it is likely that some would have exceeded this if the norm tables had been based on a wider standardization sample.

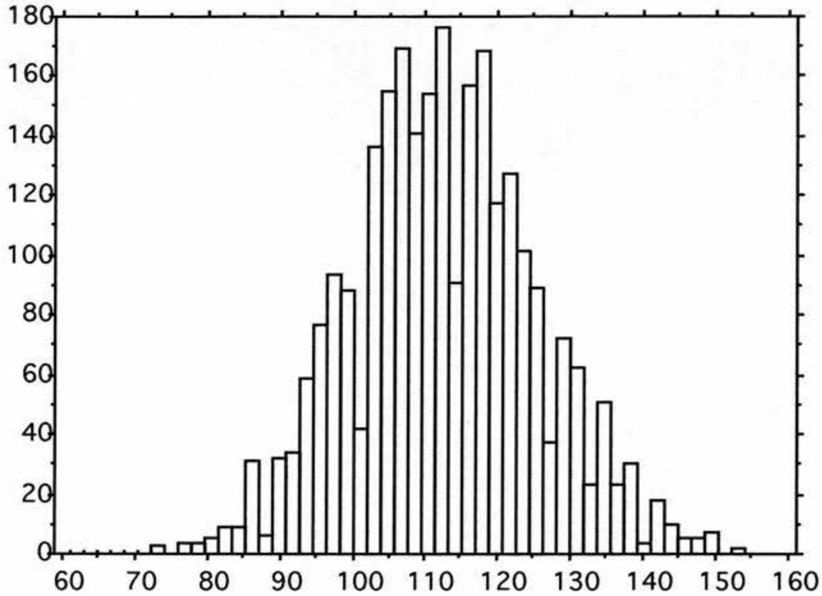
The distribution of subjects' averaged PMA IQs is illustrated in figure 4.2 (a)

Fig 4.2 (a) Distribution of Means of PMA IQs.



As explained earlier, this distribution is somewhat narrower than the distribution for the individual PMA IQs. IQs on the individual tests range from 61 (lowest score PS test) to 153 (highest score VM & SR).
The distribution of all PMA IQ scores (N= 524 IQs x 5 tests) is shown in figure 4.2 (b). [Mean sample IQ is, of course, the same in both cases]

Fig 4.2 (b) Distribution of IQs obtained by all subjects on all Five Primary Mental Abilities tests.



CHAPTER 5

STUDY 1:

AGE DIFFERENCES IN THE STRENGTH OF G.

Having established that the sample is suitable for the purposes of investigating the differentiation hypotheses, we can now turn to the central issue of study 1: *Does the influence of Spearman's g become weaker as age increases?*

In the following attempts to answer this question, raw scores will be used as the units of analysis, rather than IQs. This is desirable since raw scores are more closely indicative of the absolute ability level (MA) for subjects on a given test than are IQ scores. It is also useful in that not all of the tests have been standardised. Whilst it would be possible to re-standardise using the current samples (N >100 in all age groups) this would not offer any great advantage over the use of raw scores, at this stage.

RAW SCORE DESCRIPTIVE STATISTICS AT EACH AGE

The battery comprises ten tests, the PMA Reasoning test having been divided into its component subtests *Figure Grouping* and *Word Grouping*. Means, standard deviations and ranges of raw scores for each test, at each age level, are summarised in table 5.1, along with coefficients of skewness and kurtosis. (Latter derived from the third and fourth moments around the mean, respectively.)* Frequency histograms, showing the distributions of scores on individual tests within each of the five age groups are given in appendix 5.1.

*It should be noted that the statistical table relating to the moment coefficient of skewness treats very small departures from normality as 'significant' in samples of this size. In practice, however, a certain degree of skewness is to be expected in psychometric measures and, in mild cases, is not cause for great concern. In an investigation of 440 large-sample achievement and psychometric measures, for example, Micceri (1989) found significant deviation from normality in all samples. Across all types of measures, 71.6% showed moderate to extreme skew and amongst psychometric measures 84% were at least moderately asymmetric. A useful rule of thumb, in interpreting the given skewness coefficients, is to assume that that values of plus or minus 1.0 to 1.5 indicate a moderate degree of skewness, whereas values exceeding 1.5 or 2.0 will be obtained with more seriously skewed data (e.g. Dunlap et al, 1994).

TABLE 5.1. MEANS, STANDARD DEVIATIONS, RANGES AND ESTIMATES OF NORMALITY FOR RAW TEST SCORES

SAMPLE	TEST	MEAN	SD	RANGE	SKEWNESS	KURTOSIS
EIGHT	Raven's Matrices	34.69	8.90	42 (14-56)	-.42	-.39
EIGHT	Verbal Meaning	32.42	8.26	38 (14-52)	.21	-.61
EIGHT	Number Facility	24.73	6.31	31 (5-36)	-.77	.55
EIGHT	Spatial Relations	12.06	4.40	18 (4-22)	.33	-.83
EIGHT	Figure Grouping	17.41	3.18	14 (10-24)	-.12	-.73
EIGHT	Word Grouping	16.16	3.43	19 (6-25)	-.16	.33
EIGHT	Perceptual Speed	17.62	5.25	29 (0-29)	-.58	.59
EIGHT	Word Fluency	13.57	3.66	19 (4-23)	-.08	.31
EIGHT	Digit Span	4.69	0.97	6 (2-8)	.66	.86
EIGHT	Figures of Speech	13.87	4.30	18 (5-23)	-.11	-.60
NINE	Raven's Matrices	39.82	7.38	39 (12-51)	-1.02	1.58
NINE	Verbal Meaning	39.27	8.83	40 (17-57)	-.10	-.35
NINE	Number Facility	29.46	4.65	30 (16-46)	-.09	.97
NINE	Spatial Relations	13.57	3.92	18 (4-22)	-.15	-.43
NINE	Figure Grouping	18.90	3.06	22 (3-25)	-1.33	5.12
NINE	Word Grouping	18.37	3.13	16 (7-23)	-.72	.28
NINE	Perceptual Speed	20.45	4.84	30 (0-30)	-.59	1.59
NINE	Word Fluency	15.56	3.99	25 (4-29)	.17	.97
NINE	Digit Span	4.99	0.96	4 (3-7)	-.04	-.73
NINE	Figures of Speech	14.74	4.71	23 (4-27)	.07	.08
TEN	Raven's Matrices	43.43	7.07	37 (16-53)	-1.3	2.3
TEN	Verbal Meaning	45.92	8.10	38 (21-59)	-.95	.63
TEN	Number Facility	32.66	3.99	21 (20-41)	-.31	.28
TEN	Spatial Relations	16.21	3.52	17 (6-23)	-.41	-.16
TEN	Figure Grouping	20.09	3.21	15 (10-25)	-.78	.51
TEN	Word Grouping	19.37	3.46	17 (8-25)	-1.17	1.22
TEN	Perceptual Speed	22.41	5.15	26 (12-38)	.34	-.12
TEN	Word Fluency	17.33	4.48	22 (8-30)	.43	.10
TEN	Digit Span	5.23	0.90	4 (3-7)	.12	-.49
TEN	Figures of Speech	14.63	4.54	19 (5-24)	-.24	-.72
ELEVEN	Raven's Matrices	46.85	5.87	34 (23-57)	-1.18	2.13
ELEVEN	Verbal Meaning	51.14	6.17	29 (31-60)	-.84	.21
ELEVEN	Number Facility	34.85	4.14	25 (22-47)	-.08	.64
ELEVEN	Spatial Relations	17.37	3.90	16 (9-25)	-.26	-.59
ELEVEN	Figure Grouping	20.91	2.53	14 (11-25)	-.83	1.18
ELEVEN	Word Grouping	20.71	2.23	10 (15-25)	-.51	-.02
ELEVEN	Perceptual Speed	27.05	5.33	27 (13-40)	.04	.17
ELEVEN	Word Fluency	19.69	4.17	21(9-30)	.01	-.19
ELEVEN	Digit Span	5.80	1.04	4 (4-8)	.03	-.52
ELEVEN	Figures of Speech	16.35	3.75	19 (7-26)	-.03	-.29
TWELVE	Raven's Matrices	48.37	4.71	36 (32-58)	-.42	.54
TWELVE	Verbal Meaning	54.31	5.30	29 (31-60)	-1.86	4.26
TWELVE	Number Facility	37.97	4.17	21 (27-48)	.15	-.10
TWELVE	Spatial Relations	18.81	3.27	15 (10-25)	-.31	-.39
TWELVE	Figure Grouping	22.22	1.96	8 (17-25)	-.66	-.33
TWELVE	Word Grouping	21.95	1.97	9 (16-25)	-.68	-.06
TWELVE	Perceptual Speed	29.75	5.13	23 (16-39)	-.28	-.46
TWELVE	Word Fluency	21.68	4.36	20 (12-32)	.38	-.14
TWELVE	Digit Span	6.17	1.07	5 (4-9)	-.05	-.20
TWELVE	Figures Of Speech	17.97	4.26	20 (8-28)	-.38	-.36

As expected, raw scores on all tests increase steadily with age, the exception being *Figures of Speech,* where the mean score for nine year olds is very slightly higher than that for ten year olds.

Eight Year Olds:

In the sample of eight year olds, the distributions of scores on all tests are broadly normal, with no evidence of serious skew or curtailment. Although a high mode in the distribution for *Number Facility* does render the skewness coefficient (-.77) above the critical value (.57 for a two-tailed test), this indicates only mild negative skew. Skew is also modest, 'though significant for the *Digit Span* test (.66).

Nine Year Olds:

For nine year olds, the majority of score distributions are broadly normal, although the skewness coefficients exceed the critical value for four of the variables [RPM -1.02, FG -.133, WG -.716, PS -.586]. Negative skew in all of these cases can be explained by the presence of one or a very small number of very low scores. Higher kurtosis coefficients for these variables can be explained in the same way.

Ten Year Olds:

In the sample of ten year olds, score distributions for *Raven's Matrices, Verbal Meaning, Figure Grouping* and *Word Grouping* are characterised by mild to moderate negative skew which, as with the sample of nine year olds, can be accounted for by a very small number of low scorers. (RPM -.28, VM -.95, FG -.78, WG -.17) Distributions for the remaining six variables do not depart significantly from normality.

Eleven Year Olds:

As with the other samples, score distributions for eleven year olds are broadly normal. Significant skewness, where it occurs [RPM -1.18, VM -.84, FG -.83] can be explained by the long negative tails of these distributions.

Twelve Year Olds:

In the sample of twelve year olds, the majority of score distributions are normal or only slightly skewed [FG -.67, WG -.68]. The exception to this is the *Verbal Meaning* test [skewness = -1.86] where the shallow negative tail extends over several values and where there appears to be some curtailment at the upper end. This would suggest a ceiling effect for this test, however only a small number of subjects have achieved the maximum score and the effect is, therefore, limited.

To summarise: In all age groups, the majority of tests conform to psychometric requirements of normality. Where score distributions depart from normality it is only to a mild or moderate extent, the exception being *Verbal Meaning* in the sample of twelve year olds, which has a more serious degree of skew. It should be noted that where skew occurs it tends to be negative, resulting from shallow tails at the lower end. This is to be expected, given the nature of the school from which subjects were drawn, with the majority of pupils somewhat 'brighter' than average but also a small number of remedial pupils.

It should also be noted that score distributions tend to become smaller as age increases. This issue will be addressed in detail in a later section, since it has implications for comparing correlation sizes across age groups. At this stage, however, the normality of the data (and hence its usefulness for psychometric study) is the focus of concern.

Correlations Between Test Variables in the Five Age Groups.

Correlations between the ten tests, in each of the five age groups, are shown in tables 5.2 (a) to (e). Given below the relevant column, in each matrix, is the average of the variable's correlations with the other nine (an estimate of the test's g-saturation).

In all age groups the majority of correlations are positive, as expected, although they tend to be higher for *Ravens Matrices* and the six PMA sub-tests than for *Digit Span, Word Fluency* or *Figures of Speech.*

In eight year olds, the test which, on average, correlates most highly with the others is *Number Facility* (.48), closely followed by *Word Grouping* (.43). A similar pattern is evident for nine year olds, with the average correlation of both *Number Facility* and *Word Grouping* being .40. In eleven year olds, it is *Raven's Matrices* and *Perceptual Speed* which show the highest overall correlations (mean r =.35 and .32, respectively). *Raven's Matrices* is the most highly g-loaded test at age twelve (mean r =.34), with *Word Fluency* the next highest (mean r =.32).

The correlation matrices, are not, in themselves, particularly informative in terms of assessing whether the influence of positive manifold differs between age groups. The next section deals with methods for distilling the matrices so as to more clearly reveal the answer to this question.

Tables 5.2 a-e.
CORRELATIONS BETWEEN RAW TEST SCORES AT EACH AGE

Tb. 5.2 a. EIGHT YEAR OLDS (N = 111)

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.349	1								
NF	.473	.531	1							
SR	.517	.364	.574	1						
FG	.437	.236	.420	.409	1					
WG	.458	.535	.556	.469	.343	1				
PS	.391	.378	.541	.505	.290	.428	1			
WF	.342	.493	.604	.376	.229	.476	.419	1		
DS	.150	.128	.168	.083	-.063	.199	.202	.221	1	
FS	.244	.309	.425	.374	.270	.400	.358	.369	.146	1
MEAN	.374	.369	.477	.408	.286	.429	.390	.392	.137	.322

Tb. 5.2 b. NINE YEAR OLDS [N=114]

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.435	1								
NF	.478	.410	1							
SR	.526	.344	.375	1						
FG	.446	.390	.474	.401	1					
WG	.46	.449	.570	.401	.544	1				
PS	.314	.263	.526	.254	.304	.444	1			
WF	.173	.064	.236	.193	.051	.318	.224	1		
DS	.191	.130	.190	.063	-.006	-.007	.147	.139	1	
FS	.273	.263	.303	.300	.280	.422	.291	.391	.196	1
MEAN	.366	.305	.396	.318	.320	.400	.307	.200	.116	.302

Tb. 5.2 c TEN YEAR OLDS [N=98]

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.555	1								
NF	.437	.382	1							
SR	.450	.363	.378	1						
FG	.515	.394	.461	.460	1					
WG	.508	.503	.395	.35	.478	1				
PS	.262	.159	.509	.387	.308	.33	1			
WF	.126	.343	.359	.231	.156	.079	.307	1		
DS	.175	.152	.118	.121	.084	.101	.110	.058	1	
FS	.122	.132	.275	.133	.040	.034	.174	.310	.163	1
MEAN	.350	.331	.368	.319	.322	.309	.283	.219	.120	.154

Tb. 5.2 d. ELEVEN YEAR OLDS [N=103]

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.315	1								
NF	.460	.312	1							
SR	.439	.251	.218	1						
FG	.428	.262	.235	.337	1					
WG	.469	.294	.327	.350	.494	1				
PS	.328	.284	.458	.397	.243	.396	1			
WF	.192	.294	.362	.102	.062	.173	.309	1		
DS	.405	.287	.283	.249	.217	.175	.221	-.092	1	
FS	.133	.110	.092	.269	.121	.143	.275	.293	.139	1
MEAN	.352	.268	.305	.290	.267	.314	.324	.188	.209	.175

Tb. 5.2 e. TWELVE YEAR OLDS [N=98]

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.364	1								
NF	.415	.106	1							
SR	.363	.434	.187	1						
FG	.356	.133	.316	.298	1					
WG	.269	.270	.132	.136	.336	1				
PS	.380	.144	.517	.285	.228	.281	1			
WF	.307	.346	.340	.288	.233	.347	.453	1		
DS	.227	.191	.241	.134	.193	.083	.251	.209	1	
FS	.337	.325	.304	.224	.150	.159	.223	.396	.206	1
MEAN	.335	.257	.284	.261	.249	.224	.307	.324	.193	.258

AGE AND SPEARMAN'S G

There are several simple estimates which can be used evaluate the strength of g at different ages. All of these are essentially ways of describing the same thing - namely the amount of common variance amongst ability test scores.

i] Average Correlation (excluding diagonals at unity)

The most important estimate of common variance between test scores is the size of their intercorrelations. The strength of g in a matrix can be estimated by averaging the individual correlations, excluding the diagonals which have a value of unity.

i i] Variance accounted for by the first principal component.

Another estimate is given by the magnitude of the first and largest eigenvalue extracted in a principal components analysis. The *total* variance in the matrix is used in the computation of this value. Since this includes the correlation of each variable with itself the estimate of g-strength will be artificially inflated.

iii] Estimate of average correlation using Kaiser's method

Kaiser (1968) has suggested that the average size of the correlations in a matrix can be estimated by subtracting 1 from the largest eigenvalue and dividing the product by the number of variables minus 1. This produces a result somewhere between (i) and (ii), since some - but not all - of the common variance in the diagonals has been removed.

iv] Average of Squared Multiple Correlations of all Variables

Finally, generality of variance can be estimated by averaging the squared multiple correlations (SMC) of all the variables. The SMC is essentially an estimate of the amount of score variance for a particular test which can be predicted from all the other tests in the matrix.* The average of the SMCs is an estimate of the total variance in the matrix which is predictable from the variables. SMCs for the individual variables and mean SMC are given below:

Table 5.3 Age Differences in Tests' Squared Multiple Correlations

AGE	8	9	10	11	12
Ravens Matrices	.39	.41	.53	.46	.37
Verbal Meaning	.43	.34	.51	.25	.34
Number Facility	.60	.52	.48	.39	.38
Spatial Relations	.47	.34	.37	.31	.32
Figure Grouping	.30	.38	.46	.30	.28
Word Grouping	.46	.52	.45	.37	.25
Perceptual Speed	.39	.33	.37	.37	.41
Word Fluency	.46	.23	.31	.32	.37
Digit Span	.10	.13	.05	.27	.13
Figures of Speech	.27	.27	.17	.16	.26

The strength of Spearman's g, estimated by each of these four methods, is shown, for the five age groups, in table 5.4 Figure 5.1 illustrates these results in graphical form.

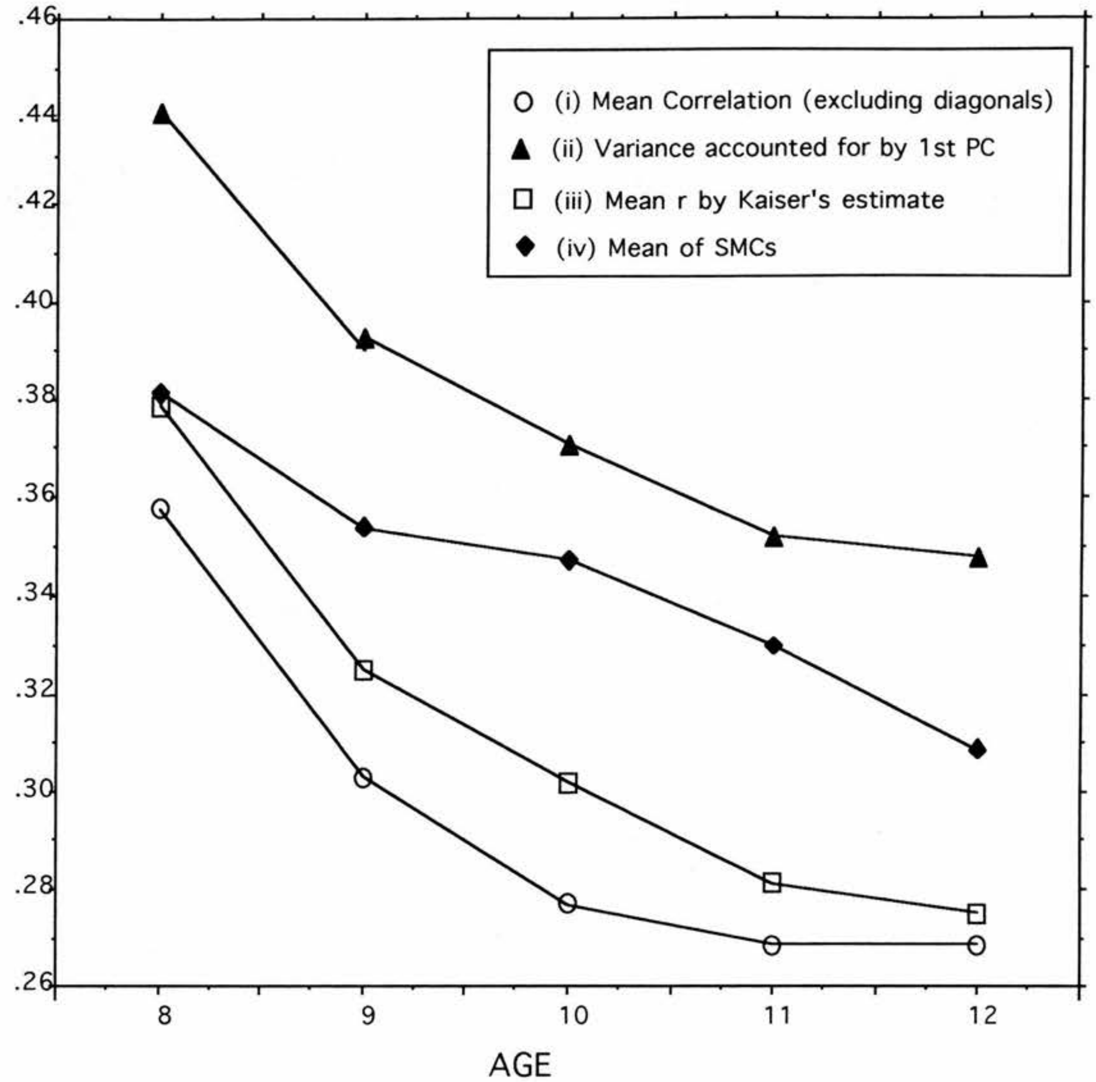
TABLE 5.4 AGE AND THE STRENGTH OF SPEARMAN'S G, ESTIMATED BY FOUR METHODS

AGE	(i) Mean Correlation (Excl. diagonals)	(ii) Variance explained by first Princ. Comp.	(iii) Mean r (by Kaiser's method)	(iv) Mean of SMCs
8	.358	.441	.379	.381
9	.303	.393	.325	.354
10	.277	.371	.302	.347
11	.269	.352	.281	.333
12	.269	.348	.275	.309

* [More specifically, it is an estimate of the percentage of variation in scores on one particular variable in the matrix which is predictable in a linear regression equation using all the other variables in the matrix. It is often used as an initial estimate of communality, prior to factor analysis, although this practice is not favoured by all psychometricians (eg. Kline, 1993).]

AGE AND STRENGTH OF SPEARMAN'S *G* BY FOUR ESTIMATES

Figure 5.1



Numbers on the *y* axis relate to correlation coefficients for three of the variables. Variance accounted for by the first principal component is expressed as a fraction of 1.00.

As can be seen from table 5.4 and figure 5.1, there is a gradual tendency for the influence of positive manifold to decrease as the age of the sample increases. The estimate which shows the most systematic decrease in the strength of Spearman's g with age, is the mean correlation. The values for the variance accounted for by the first principal component and mean r by Kaiser's method, do not differ between ages nine and ten, and the average of the squared multiple correlations increases across the same interval. Nevertheless, all four methods of estimation reveal a clear decrease in g-strength over the age range eight to twelve years.

Spearman rank order correlations were computed for age against all four estimates (table 5.5). In all cases, the relationship is significant at the five percent level, using a one tailed test.

Table 5.5

Spearman Rank Order Correlations between Age and Four Estimates of g-Strength

Age & mean correlation	Rho = -1.00 p<.05
Age & % var. explained by Ist PC	Rho = -.97 p<.05
Age & mean r by Kaiser's method	Rho = -.97 p<.05
Age & mean of SMCs	Rho = -.90 p<.05

Number of Factors at Each Age

As discussed in chapter 2, an important source of evidence for those who have researched the differentiation hypothesis in the past, has been the number of factors extracted from matrices of ability test correlations at different ages. The less pervasive Spearman's g, the reasoning goes, the more factors will be revealed when scores on a battery of tests of diverse abilities are subjected to factor analysis. This approach has also been favoured by the most recent psychologist to re-analyse the early age differentiation studies (Carroll, 1993). The usefulness of this approach, for the purposes of present study, will now be considered.

Comparing numbers of factors at each age is meaningless unless the method of estimation is stated (cf. Cattell, 1978). Unfortunately this practice has not always been followed in past studies.

Three methods of factor extraction will be examined in the following age comparisons: eigenvalues greater than 1, root curve analysis and the 75% variance rule.

1) Kaiser-Guttman Rule: Eigenvalues greater than 1 (Kaiser, 1960)

This is the most popular extraction rule for data from a principal components analysis. In essence it dictates than only factors which explain at least ten percent of the variance in the matrix should be accepted. Despite its wide use, the technique has been criticized by many as misleading (e.g. Carroll, 1993; Kline; 1993). The problem is that it tends to underestimate the number of factors when variables are substantially correlated and to overestimate the number when the average correlation between variables is low. The former is particularly problematic in analyses of ability test data, since the influence of

Spearman's g is strong. To elaborate: in principal components analysis the largest eigenvalue reflects the variance in the matrix which is common to all the variables. When - as with ability tests- this comprises most of the inter-test variance, very little variance will be left over after it has been extracted and much of what *is* left will be specific to particular tests.

2) Root Curve Method/Scree Test (Cattell, 1966)

The number of factors extracted by this rule is traditionally determined by examining the plot of eigenvalue magnitude against eigenvalue number. This distribution is referred as the 'root curve'; reflecting its characteristic shape. Beyond the 'elbow' of the root curve, the differences between roots become smaller and the values begin to approach zero. At some point both the differences and the amount of variance explained, become negligible. Values beyond this point are referred to as the 'scree' and are discarded. The root curve method used in the current analysis is essentially Cattell's scree test, as performed by the *Statview* factor analysis program for the Apple Macintosh.

3) 75 % of the variance

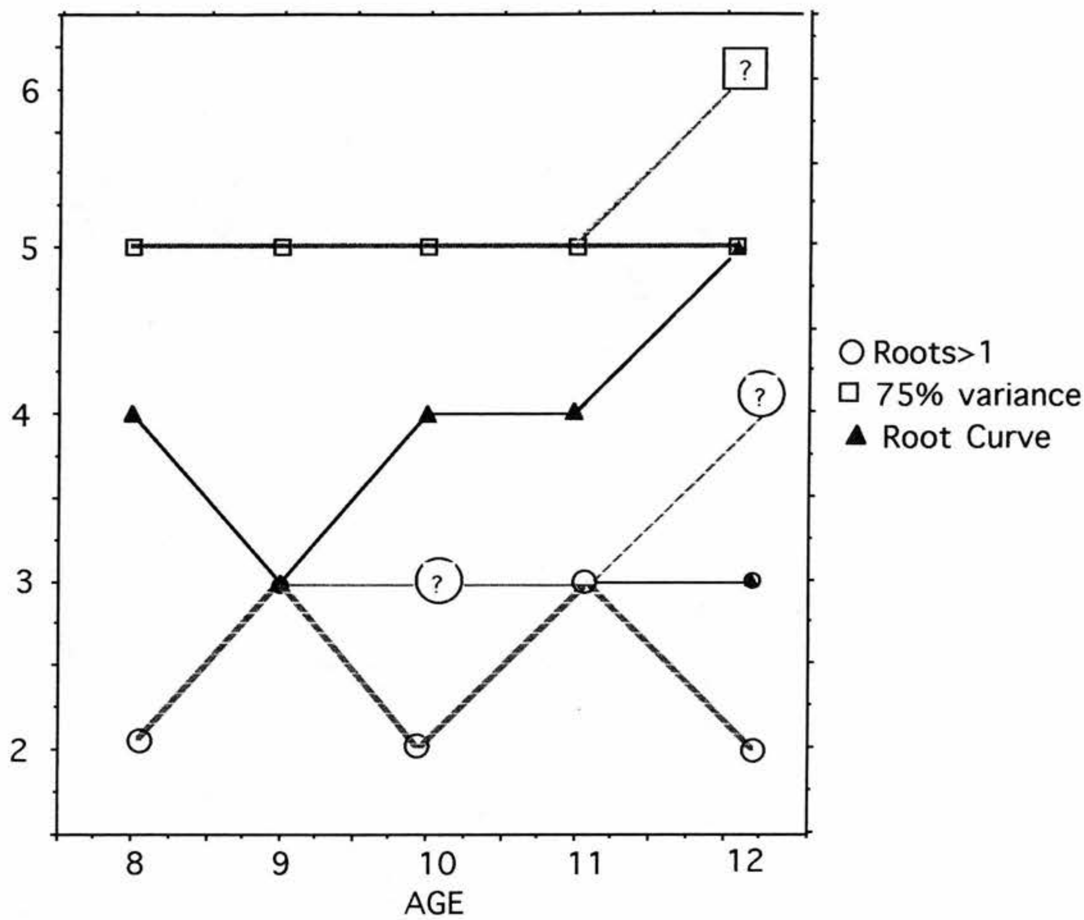
A third rule is, simply, that no further factors should be extracted beyond the point where the cumulative variance explained by the eigenvalues exceeds 75%.

The number of factors which would be extracted using each of these criteria is shown in figure 5.2 and table 5.6.

* [There are several other methods of determining the number of useful factors from a list of eigenvalues. These are mostly used in 'common factors' analyses, where a figure other than unity has been placed in the diagonal of the correlation matrix. (e.g. Montanelli & Humphreys, 1976)]

AGE AND NUMBER OF FACTORS EXTRACTED USING THREE METHODS

Figure 5.2



Question marks indicate numbers of factors revealed when criteria for acceptance are relaxed only slightly (see notes below table 5.6).

Table 5.6

AGE	Roots > 1	75 % Variance	Root Curve
EIGHT	2	5	4
NINE	3	5	3
TEN	2 (3)*	5	4
ELEVEN	3	5	4
TWELVE	2 (3)	5 (6)*	5

*Age 10 EV3 = .999
Age 12 EV3 = .994

*Age 12: Five factors account for 73.3% of variance. 6 factors account for 81.0%

Using the Kaiser-Guttman rule, there would, initially, appear to be no consistent relationship between age and number of factors, with two factors appearing at ages eight, ten and twelve and three at ages nine and eleven.* Closer inspection of the eigenvalues, however, reveals that the third components at ages ten and twelve have values of .999 and .994, respectively, and amongst twelve year olds, the fourth eigenvalue (.951) also explains close to ten percent of the variance. The extraction of two, rather than three, factors is therefore based on very minor differences in these groups. If the figures exceeding .95 are rounded-up to 1.00 this produces a distribution like that shown by the dotted line in figure 5.2, with two factors at age eight, three at ages nine, ten, and eleven, and four at age twelve. Although such differences to not indicate a dramatic quantitative or qualitative shift in g-strength with age, they do suggest a gradual weakening.

The age differences revealed using the root curve method do not follow a consistent trend, although the number of factors increases from three to five between the ages of nine and twelve. Decisions about number of factors using this rule are, again, not based on dramatic differences between ages. As with the roots>1 rule, the root curve method cannot be relied on to provide a definitive guide to the number of factors worth extracting for later factor analysis. Despite these problems, the two methods are useful as general indicators of differentiation and they provide further evidence for the weakening of g with age.

In terms of assessing the strength of positive manifold at different ages, the 'numbers of factors' approach offers no great advantage over the other methods described and can, indeed, be misleading. For this reason, number of factors will not be used as a criterion for testing the hypothesis.

*For the reasons already stated, the first of these factors can, in every case, be assumed to be Spearman's g , with the next one or more factors accounting for the remaining variance.

The data thus far presented indicates a clear tendency for the influence of positive manifold to decrease as age increases. Although the differences between age groups are moderate, they are of the order observed in previous investigations. It is not possible to draw firm conclusions, however, without first addressing an important an issue which has been overlooked in many previous studies on age differentiation, namely, the possibility that the results are artifactual.

CORRECTING FOR STATISTICAL ARTIFACT

There are two key sources of potential bias in the correlations obtained so far which must be considered before conclusions about true score differentiation can be drawn. These are *test unreliability* and *range restriction.*

TEST RELIABILITY

The extent to which a test can correlate with other tests (and hence its validity) will be constrained by the extent to which it correlates with itself, ie. its reliability. It is therefore important to establish the degree to which each test in the battery is reliable within each age group before drawing conclusions about differentiation. If age differences in test reliabilities exist, it will also be necessary to correct the obtained correlations for these effects.

CHOICE OF RELIABILITY ESTIMATE

There are three main methods of estimating test reliability. The first of these, *test-retest reliability*, is concerned with the stability of test scores over time. By this method subjects are tested with the same instrument on two separate occasions and the results correlated. The test-retest method is problematic for two reasons: If the time between testings is too short, performance on the second occasion will be subject to practice effects (e.g. memory for and rehearsal of test material, improved test-taking skills). The longer the time period between testings, however, the more opportunity there is for other sources of error to operate. In the case of school children - as in the present study - this might involve an increase in learned skills and also, perhaps, in ability (Mental Age). The test-retest method did not represent an option in the present study, since subjects took each test on only one occasion.

The second method, *parallel forms reliability*, involves the testing of subjects with separate versions of the same test, usually on the same occasion. This overcomes many of the problems of the test-retest method, however it is still prone to practice effects and other sources of error associated with repetition (e.g. boredom).

The third method, *internal consistency reliability*, involves correlating one or more parts of the test with the other part(s). In the *split-half* method these are the two halves of the test. Since many tests are designed in such a way that the difficulty level of the items progressively increases, the two test halves are generally the odd and even items, rather than the first and second halves of the test. This method has many advantages over the test-retest and parallel form methods, such as the fact that testing takes place on the same occasion and the test does not have to be repeated. The main problem associated with the split-half method is that it effectively assesses the reliability of only half the test. The split-half correlations can, however, be corrected for the full test length using the Spearman-Brown prophecy formula.

Another estimate of a test's internal consistency reliability is Chronbach's alpha (α). This is derived from the correlation of each item in a test with the other items. While this method of reliability estimation may be somewhat more accurate than the split-half method, in this study the differences were not thought to be so substantial as to justify the much greater investment of time needed to separately code over 15000 individual items.

Estimating Test Reliabilities

For the reasons mentioned, it was decided to estimate test reliability on the basis of internal consistency, using the split-half method. (Except in the case of Digit Span, see below.)

The calculation of reliability coefficients necessitated the re-marking of all raw data. For eight of the tests separate totals were calculated for odd and even items. Since the *Word Fluency* test does not contain 'items' as such, the odd/even split was based on alternate responses.

In the case of *Digit Span* the 'split half' method was inappropriate since, unlike the other tests in the battery, obtained score did not relate to number of items correct, but rather to maximum set size recalled. To recap, three digit strings were presented for each set size and the relevant 'span' was said to be achieved when one or more of these strings was reproduced correctly. It was decided, therefore, to adopt an approach approximating the parallel forms method in order to assess test reliability. For this purpose, two 'scores' were coded for each subject, the first of these (labelled 'x') being the size of the longest digit string recalled correctly and the other (labelled 'y') as the size of the last correctly reproduced string.

Reliability estimates for each test within the whole sample and the individual age groups are shown in table 5.7. With the exception of Digit Span*, all split-half correlations have been adjusted for the full test-length using the Spearman-Brown formula.

*Since Digit Span test was not split into halves in order to calculate the reliability, the x/y correlation can be taken as the reliability estimate.

TABLE 5.7
TEST RELIABILITIES WITHIN WHOLE SAMPLE AND INDIVIDUAL AGE GROUPS

TEST	FULL SAMPLE [N = 527]	AGE 8 [N = 111]	AGE 9 [N = 114]	AGE 10 [N = 100]	AGE 11 [N = 102]	AGE 12 [N = 100]	Average Across Age Groups
Raven's Matrices	.922	.925	.890	.900	.790	.815	.864
Verbal Meaning	.946	.888	.883	.921	.886	.843	.884
Number Facility	.959	.929	.907	.907	.925	.960	.926
Spatial Relations	.828	.783	.766	.667	.800	.740	.751
Figure Grouping	.742	.598	.685	.812	.636	.484	.643
Word Grouping	.780	.660	.689	.758	.654	.570	.666
Perceptual Speed	.956	.914	.912	.934	.940	.928	.926
Word Fluency	.962	.955	.960	.917	.945	.943	.944
Digit Span	.858	.851	.769	.794	.849	.826	.818
Figures of Speech	.813	.763	.788	.799	.765	.850	.793
Mean of reliabilities	.877	.827	.825	.841	.819	.796	

N.B. Reliability coefficients are based on split-half (odd/even) correlations corrected for the full test length using the Spearman-Brown formula. The exception to this is the Digit Span test for which reliability estimates are based on the correlation between parallel forms.

As shown in table 5.7 the mean reliability of tests in the battery (bottom row) exceeds 0.8 in all age groups* and in the sample as whole. Reliabilities in the full sample are, in general, higher than in the individual age groups, since the number of subjects and the distribution of scores are larger.

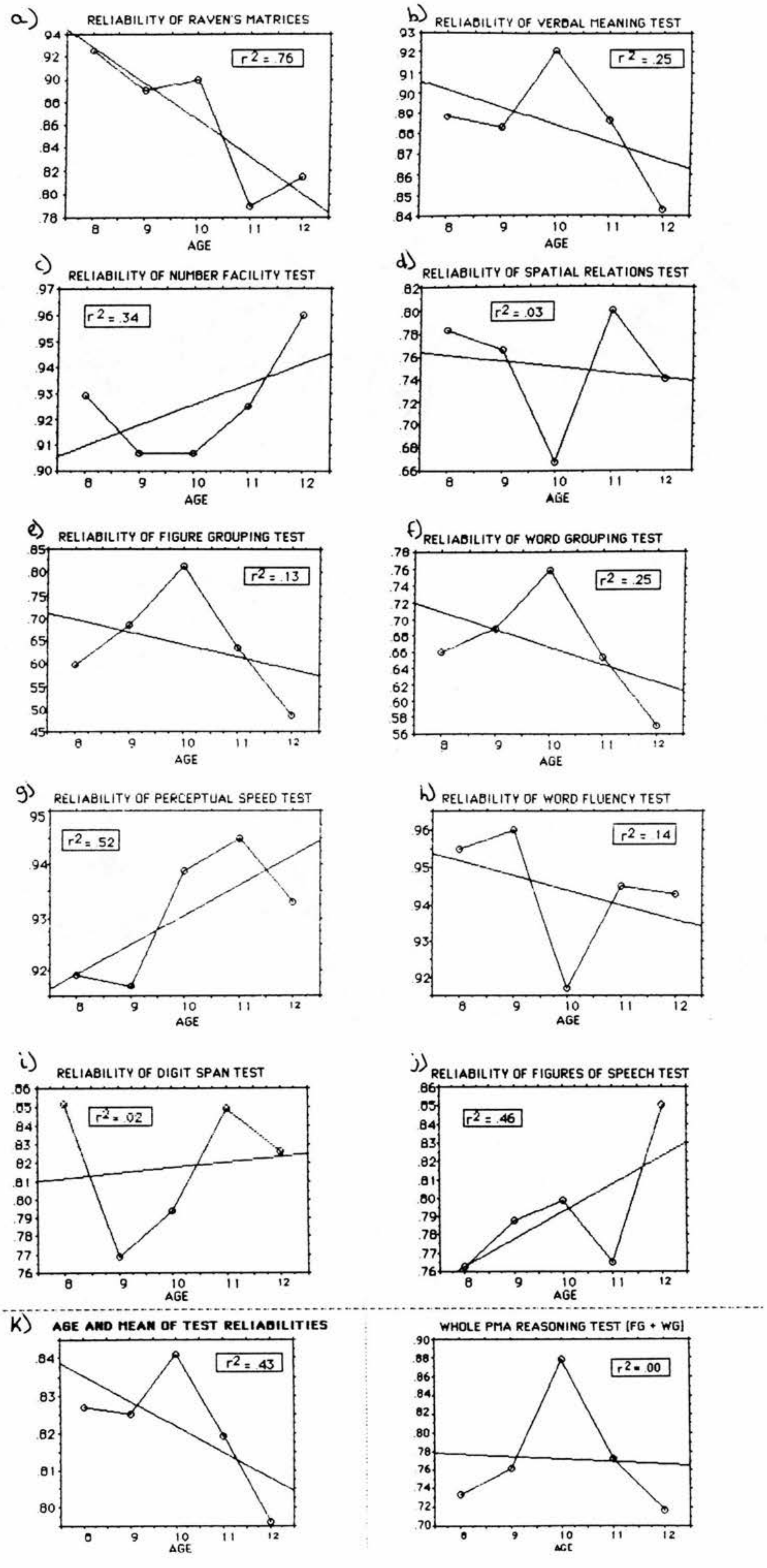
When the individual test reliabilities obtained at each of the five ages are averaged (column to the far right) reliabilities exceed 0.8 for six out of the ten tests (RPM .86, VM .88, NF .93, PS .93, WF .94 DS .82). The average reliability of the *Figures of Speech* test (.79) is high for a 'creativity' test, for which figures below .50 are more commonly obtained (Rust & Golombok, 1989). The other tests with average reliabilities below .8 (*Spatial Relations*, .75; *Figure Grouping*, .64; *Word Grouping*, .67) have fewer items than the other tests (25 vs 30-60), and they have odd numbers of items. It is therefore to be expected that their reliabilities are somewhat lower. The two tests with the lowest reliabilities (FG and WG) were initially designed as part of a global Reasoning test, for which the overall reliability (.77) is understandably higher. The average reliability of *Digit Span* (.82) is also high given its small number of 'items'. This can be explained by the homogeneity of the two 'items' which were correlated.

Age differences in the reliabilities of individual tests across ages and in the sample as a whole are shown in figures 5.3 (a) to (k).

*(when rounded to two decimal places)

FIGURES 5.3 a) to k)

AGE AND TEST RELIABILITIES



Figures 5.3 (a) to (k) illustrate the reliabilities of each test and the whole battery at each age. Regression lines have been drawn through the scatter plots for illustrative purposes and coefficients of regression and correlation calculated. The latter are summarised in table 5.8.*

Table 5.8. Pearson Correlations of Age with Test Reliability (df=3)

TEST	r Age & Reliability	TEST	r Age & Reliability
Raven's Matrices	-.869	Perceptual Speed	+.721
Verbal Meaning	-.496	Word Fluency	-.370
Number Facility	+.582	Digit Span	+.134
Spatial Relations	-.158	Figures of Speech	+.676
Figure Grouping	-.365		
Word Grouping	-.501	Mean of whole battery	-.654

The above correlations, indicate that reliability increases between the ages of eight and twelve for four of the ten tests in the battery, and decreases for six. The average reliability of the battery decreases with age. Many of the coefficients are only moderate in size, however, and all are based on a small number of observations (5). It is therefore essential that the scatterplots are examined more closely before drawing conclusions about the consistency of the change in reliabilities with age.

The first variable, *Raven's Matrices*, shows the most pronounced drop in reliability with age (r =-.869). This is somewhat surprising given that this test is generally known to have high reliability and is well standardised across the age range tested. The age differences in reliabilities are reasonably small, however, with the figures for eight and twelve year olds being .93 and .82, respectively. The patterns of reliability for *Verbal Meaning, Figure Grouping* and *Word Grouping,* resemble an inverted U. In all cases, this is accounted for by higher

*The appropriateness of computing such coefficients may be questioned, given that the relationships shown in figures 5.3 a to k are non linear. Nevertheless, they are useful in that they give a broad indication of the degree to which differences in test reliabilities might influence the general pattern of age differences in the strength of g.

reliability in the group of ten year olds, although this does not change the general (negative) direction of the relationship. The reliabilities of *Number Facility* show a U-shaped relationship with age. This finding is common in short tests covering a large age range, since at the youngest and oldest ages some items may, respectively, be too difficult or easy for a proportion of subjects and this will increase the reliability coefficient. Nevertheless the differences are small (min .91 max .96).

There does not appear to be a consistent relationship between reliability and age for the *Spatial Relations* test, although it should be noted that reliability drops below .70 in the sample of 10 year olds. Reliability of *Digit Span* shows no clear relationship to age. With the exception of the sample of eleven year olds, the *Figures of Speech* test shows a systematic increase in reliability with age. Such an increase could be accounted for by reduced measurement error in older groups, as might be caused by their greater ability to understand the test instructions.

When the figures for the full battery are averaged, the relationship between reliability and age is negative (r= -.65), although the age differences are small and are mostly accounted for by the reliabilities of the ten and twelve year old samples (.84 vs .80).

Using the reliability coefficient and standard deviation of each test, it is possible to estimate the extent to which obtained scores deviate from 'true' scores. This is achieved by calculating the standard error of measurement.

Formula for calculating the standard error of measurement.

$$\text{S.E.} = \text{sd}\sqrt{1-\text{reliability}}$$

Multiplying the standard error of measurement by 1.96 gives the 95% confidence intervals for each test score. These figures are summarised in table 5.9.

TABLE 5.9 STANDARD ERRORS OF MEASUREMENT AND 95% CONFIDENCE INTERVALS FOR EACH TEST ACROSS AGE GROUPS. [DERIVED FROM RELIABILITY ESTIMATES]

AGE	RAVEN'S MATRICES		VERBAL MEANING		NUMBER FACILITY		SPATIAL RELATIONS		FIGURE GROUPING		WORD GROUPING		PERCEPTUAL SPEED		WORD FLUENCY		DIGIT SPAN		FIGURES OF SPEECH	
	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.	S.E.	95% conf.
EIGHT	2.43	4.77	2.77	5.43	1.68	3.29	2.05	4.02	2.01	3.94	2.00	3.92	1.54	3.02	0.78	1.53	0.38	0.74	2.09	4.10
NINE	2.45	4.81	3.02	5.93	1.42	2.78	1.90	3.72	1.72	3.37	1.75	3.42	1.44	2.82	0.80	1.56	0.46	0.90	2.17	4.25
TEN	2.23	4.37	2.27	4.46	1.22	2.39	2.03	3.98	1.39	2.73	1.70	3.34	1.32	2.59	1.21	2.37	0.41	0.80	2.04	3.99
ELEVEN	2.69	5.27	2.08	4.08	1.14	2.23	1.74	3.41	1.52	2.99	1.31	2.57	1.31	2.57	0.98	1.91	0.41	0.80	1.82	3.56
TWELVE	2.03	3.97	2.10	4.12	0.83	1.63	1.67	3.27	1.41	2.76	1.29	2.53	1.38	2.71	1.04	2.04	0.45	0.88	1.65	3.24

Correcting for the Effects of Differential Test Reliability

As noted previously, differences in the reliabilities of individual tests will obscure the true pattern of correlations at different ages. This is of particular concern given the negative relationship between age and average reliability for the whole test battery, since this could account for the reduction in g-strength observed in the raw data.

To remove the effects of differential test reliabilities, the following formula was applied to the correlations obtained at each age.

Formula to correct correlations for the effects of attenuation due to test unreliability.

$$rxy\ (\text{corrected}) = \frac{rxy\ (\text{obtained})}{\sqrt{rxx}\ \sqrt{ryy}}$$

Where:

rxy (corrected) = correlation between tests x and y adjusted for the effects of unreliability

rxy (obtained) = the obtained correlation between tests x and y.

rxx = reliability of test x

ryy = reliability of test y.

The above formula adjusts each correlation coefficient to the level which would be expected if both tests were perfectly reliable. The process thus removes the contaminating effect of age differences in test reliabilities, allowing a better picture of true g-strength to emerge.

Adjusted matrices are given in appendix 5.2. In each case, figures are also supplied for the average correlation of each test with all others and for the overall average of the matrix (excluding diagonals).

As with the uncorrected matrices, estimates of Spearman's g at each age were recalculated following correction for the effects of test unreliability. It was decided, however, to concentrate on the two key measures of *g*, mean correlation excluding diagonals and percentage of variance accounted for by the first principal component.[1] Values of unity were replaced in the diagonals of each matrix before calculating principal components. (When applied to the diagonals the correcting procedure raises the values above 1.)

Age differences in estimates of g, following correction for unreliability, are illustrated in figure 5.4 (a) and table 5.10 (bold type). Both methods reveal a decline in g-strength with age, similar to that found with the uncorrected data. This decline is more pronounced when estimated as the variance explained by the first principal component, with the average correlation being approximately the same for eleven and twelve year olds.

Estimates derived from corrected matrices are compared to those obtained *before* correction in figures 5.4 (b) and (c) and table 5.10. From these it would appear that whilst correcting for test unreliability produces a general increase in amount of inter-test correlation, it does not change the overall pattern of differentiation with age.

[1] Kaiser's method of estimating mean r produces very similar results and little information is added by its inclusion. Estimates based on squared multiple correlations are distorted following the application of the correcting procedure, with the low reliability of Figure Grouping and Word Grouping at age twelve artificially inflating the mean of SMCs in this age group.

AGE AND STRENGTH OF SPEARMAN'S *G* FOLLOWING CORRECTION FOR ATTENUATION DUE TO TEST UNRELIABILITY.

Figure 5.4 a. Mean Correlation and Percentage of Variance Explained by the First Principal Component Across Five Age Groups.

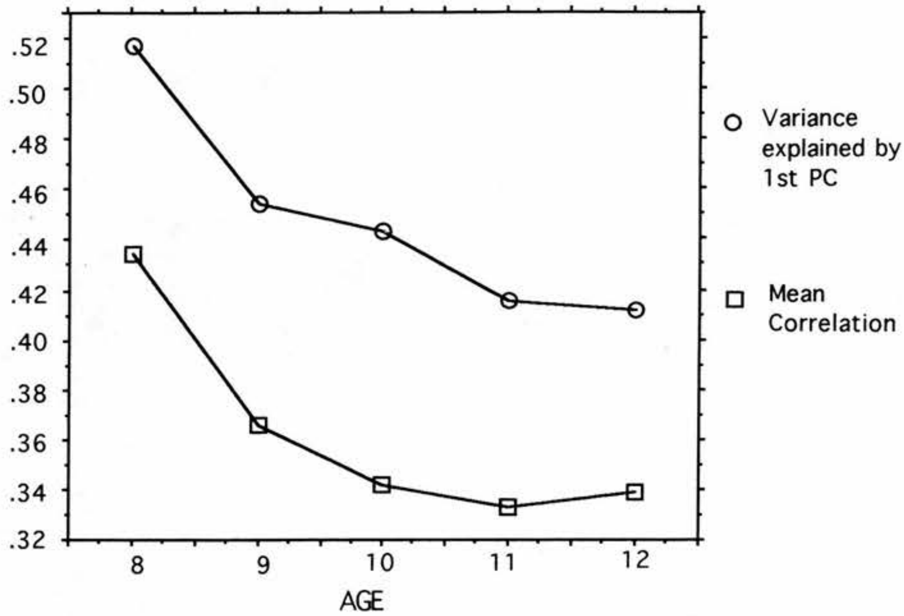


Fig. 5.4 b. Age and Average Correlation Before & After Correction for Unreliability

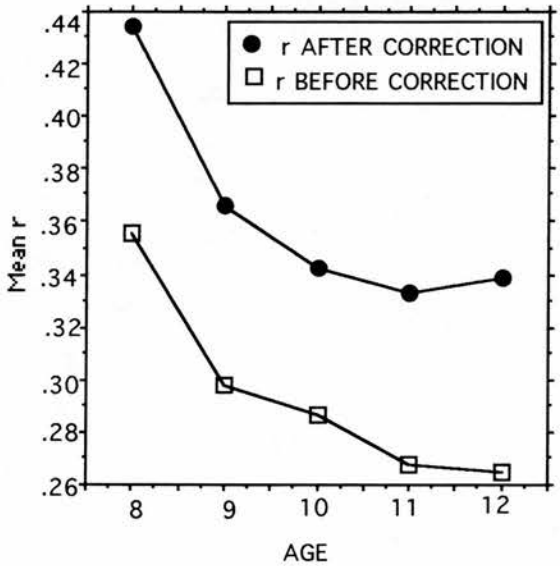


Fig. 5.4 c. Age and Variance Explained by 1st PC Before & After Correction for Unreliability

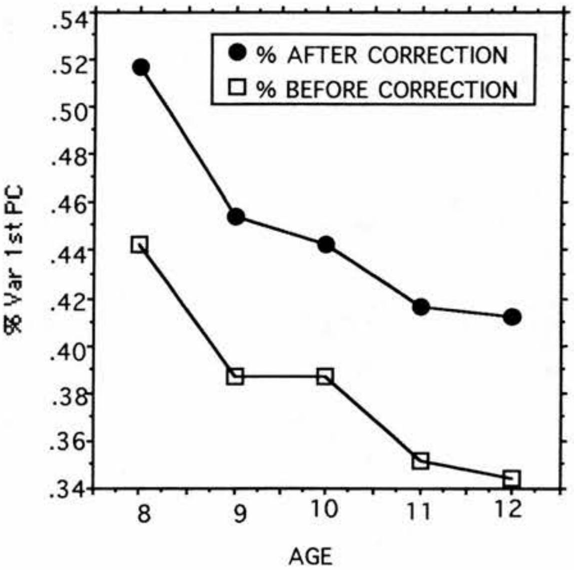


Table 5.10. Mean Correlation and Percentage of Variance Explained by the First Principal Component Before and After Correction for the Effects of Test Unreliability.

AGE	Mean correlation (exc. diagonals) AFTER CORRECTION	Mean correlation (exc. diagonals) BEFORE CORRECTION	Variance accounted for by first PC AFTER CORRECTION	Variance accounted for by first PC BEFORE CORRECTION
EIGHT	.434	.356	.517	.443
NINE	.366	.298	.454	.387
TEN	.342	.287	.443	.387
ELEVEN	.333	.268	.416	.352
TWELVE	.339	.265	.412	.344

The relationship between age and g-strength, before and after correcting for unreliability, is summarised below:

Table 5.11 Pearson correlations expressing the relationship between age and g-strength before and after correcting for the effects of test unreliability.

	BEFORE CORRECTION	AFTER CORRECTION
Age & average correlation:	-.91	-.85
Age & variance accounted for by first principal component:	-.94	-.93

Table 5.11 confirms that the corrections have had little effect on the negative relationship between age and g, although the strength of this relationship is lowered very slightly.

To summarise; the process of correcting correlations for the contaminating effects of test unreliability does not radically alter the pattern of reduced g-strength with age found in the original matrices.[1] Although the apparent systematic decrease in average test reliability with age might have led to the expectation of such a change, it is perhaps unsurprising in view of the small differences between age groups (min. average reliability = .72 max. = .86). The modest effects of such differences on actual test scores (and, consequently, on the correlations between them) can be illustrated by considering the test with the most pronounced negative relationship between reliability and age - *Raven's Matrices*. The difference in RPM reliabilities between the twelve and eight year old samples (.93 vs .82) is equivalent to a difference in measurement error (due to unreliability) of only plus or minus 0.8 of an item.[2] In contrast, the two age groups differ in range by 16 score points and in s.d. by 4.2 .

[1] although it does, to varying extents, raise the values of the correlation coefficients.

[2] 95% confidence intervals derived from the Standard Error of Measurement.

Differential test reliability across age-groups is, however, only one source of statistical bias likely to influence the magnitude of observed correlations. If the apparent differentiation with age is indeed artifactual, then it must have an additional, more powerful, source. In this regard, the most likely candidate is the extent to which the samples are of equivalent heterogeneity with respect to ability. This is the focus of the next section.

RANGE RESTRICTION EFFECTS

It is well known that restriction of range in one or more test variables, will lower the correlations between them.[1] For this reason it is essential that we consider the *range* of scores in the distributions for each test, and particularly their *dispersion around the mean*, as shown by the standard deviation.[2]

AGE AND SCORE DISPERSION

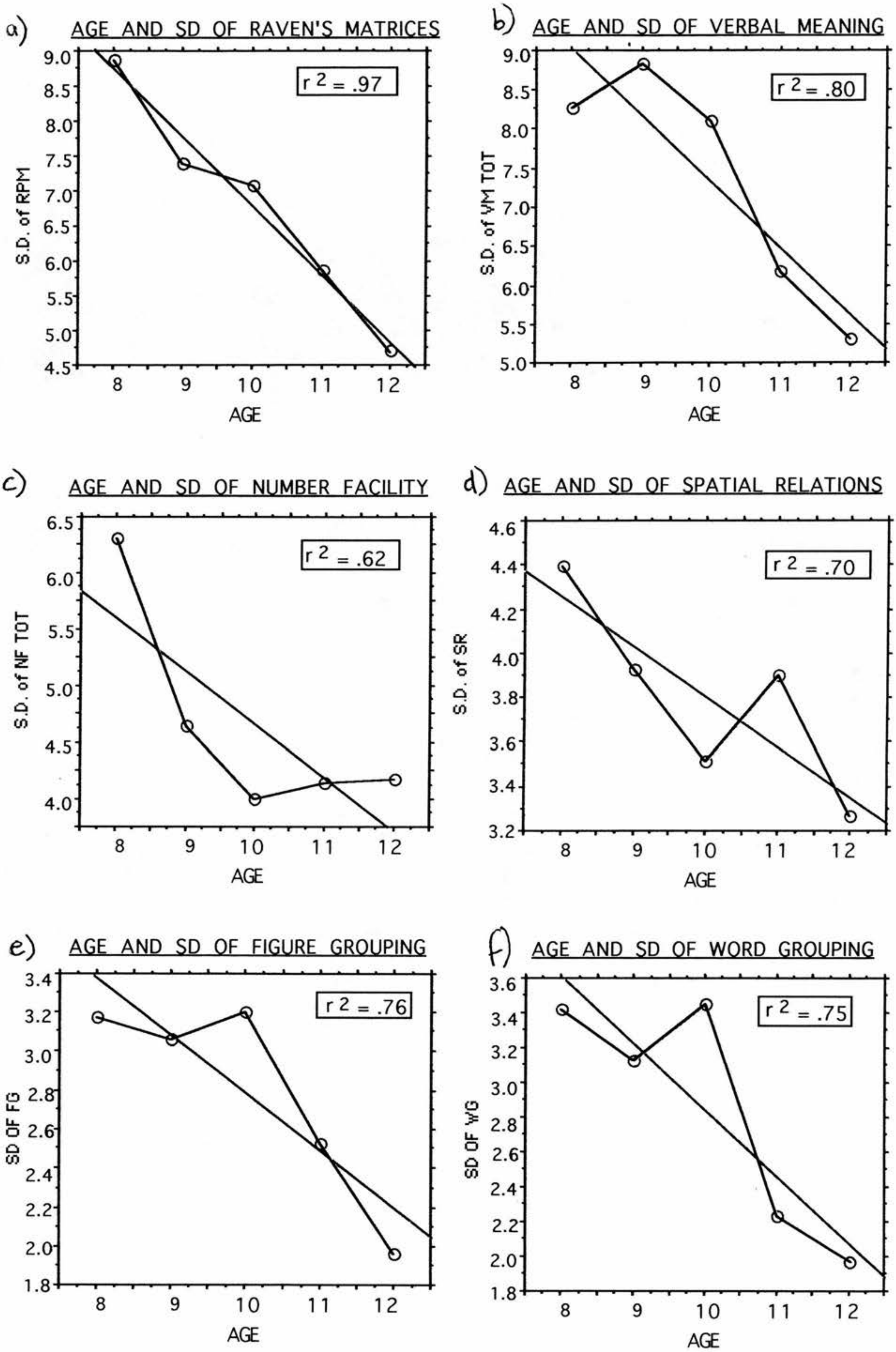
As noted earlier, although most of the test variables have a broadly normal distribution, several tend to become progressively narrower in samples of increasing chronological age.

Standard deviations and ranges of each test in the five age groups are shown in table 5.12. The relationship between age and dispersion, for each test and for the battery as a whole, is illustrated in figures 5.5 (a) to (k).

1 This has more to do with restriction in the *standard deviation* of scores than in their *range* per se, because it is the SD which is the statistical feature used in the computation of the correlation coefficient.

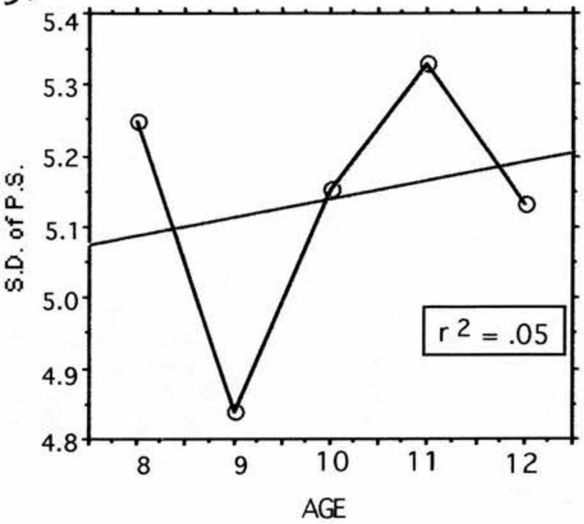
2 SD and range will be related perfectly in an ideal normal distribution. Most test score distributions deviate slightly from the ideal, however. Take, for example, two distributions of test scores, which are equal in total range. One is characterised by a tall, narrow centre with long, shallow tails extending over a many values. For the other, the sides of the bell-shape descend sharply, close to the maximum and minimum scores. The SD of the latter distribution will be much larger, even though range is the same.

AGE AND STANDARD DEVIATION OF RAW SCORE DISTRIBUTIONS

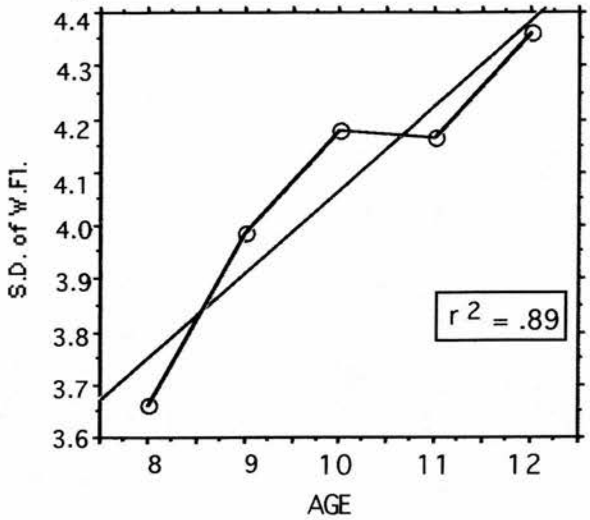


AGE AND STANDARD DEVIATION OF RAW SCORE DISTRIBUTIONS

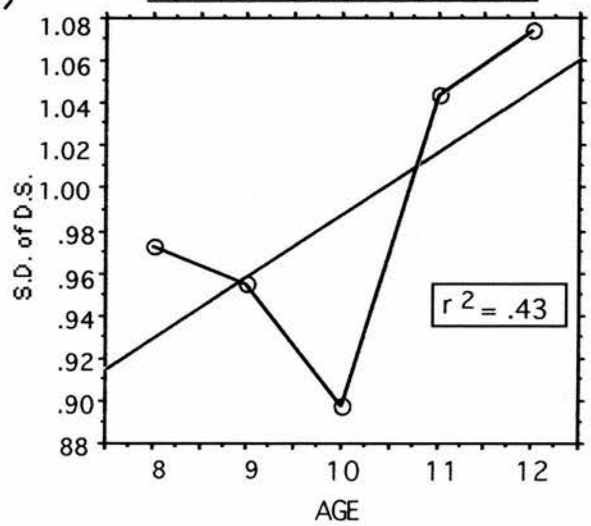
g) AGE AND SD OF PERCEPTUAL SPEED



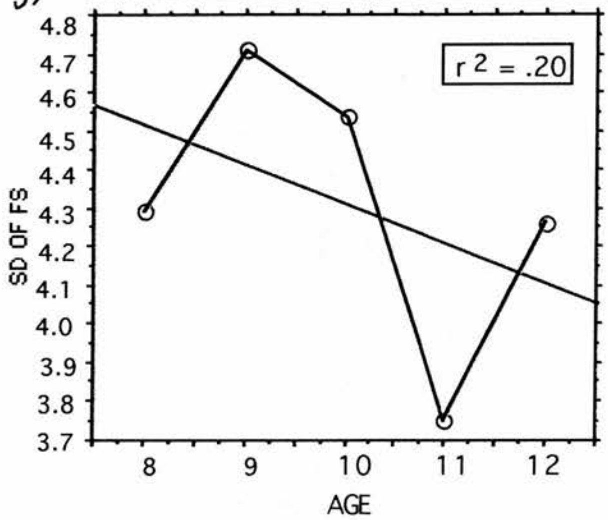
h) AGE AND SD OF WORD FLUENCY



i) AGE AND SD OF DIGIT SPAN



j) AGE AND SD OF FIGURES OF SPEECH



k) AGE AND MEAN OF STANDARD DEVIATIONS FOR ALL TEN TESTS

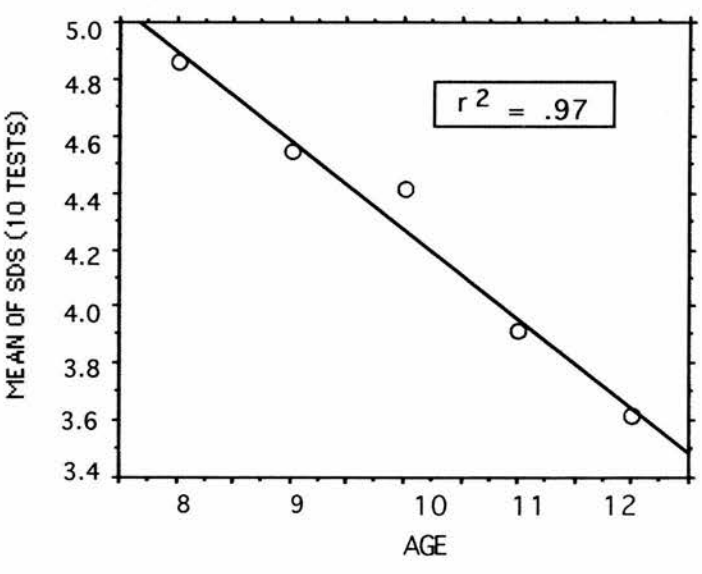


Table 5.12 Score Distributions (Standard Deviation, Range) In Each Age Group

	EIGHT		NINE		TEN		ELEVEN		TWELVE	
	SD	Range	SD	Range	SD	Range	SD	Range	SD	Range
Raven's Matrices	8.87	42	7.38	39	7.07	37	5.87	34	4.71	26
Verbal Meaning	8.26	38	8.83	40	8.10	38	6.17	29	5.30	29
Number Facility	6.31	31	4.65	30	3.99	21	4.14	25	4.17	21
Spatial Relations	4.40	18	3.92	18	3.52	17	3.90	16	3.27	15
Figure Grouping	3.18	14	3.06	22	3.21	15	2.53	14	1.96	8
Word Grouping	3.43	19	3.13	16	3.46	17	2.23	10	1.97	9
Perceptual Speed	5.25	29	4.84	30	5.15	26	5.33	27	5.13	23
Word Fluency	3.66	19	3.99	25	4.18	22	4.17	21	4.36	20
Digit Span	0.97	6	0.96	4	0.90	4	1.04	4	1.07	5
Figures of Speech	4.30	18	4.71	23	4.54	19	3.75	19	4.26	20
AVERAGE	4.86	23.4	4.55	24.7	4.41	21.6	3.91	19.9	3.62	17.6

Pearson correlations, indicating the degree to which age and standard deviation are related, are given in table 5.13.

Table 5.13 Pearson Correlations Between Test Standard Deviation and Sample Age*

Test	Correlation between SD and age	
Raven's Matrices	-.987	[p< .01]
Verbal Meaning	-.896	[p< .05]
Number Facility	-.787	[n.s.]
Spatial Relations	-.837	[p< .10]
Figure Grouping	-.873	[p< .10]
Word Grouping	-.866	[p< .10]
Perceptual Speed	+.217	[n.s.]
Word Fluency	+.945	[p< .05]
Digit Span	+.653	[n.s.]
Figures of Speech	-.448	[n.s.]
MEAN OF S.D.s	-.988	[p< .05]

*D.F.=3. Given significance levels are for a two-tailed test.

As can be seen from the preceding figures and tables, score distribution declines with increasing age for seven of the ten tests in the battery. The negative relationship between age and standard deviation is significant to the 10% level, or better, in five of these cases (*Raven's Matrices* -.99, *Verbal Meaning* -.90, *Spatial Relations* -.84, *Figure Grouping* -.87, *Word Grouping* -.87). The figure for *Number Facility* (-.79) is close to the critical value, however for *Figures of Speech* (r = -.45) the relationship between age and s.d. is inconsistent [see fig. 5.5 (j)] .

Interestingly, score standard deviation increases with age for three of the tests; *Perceptual Speed* (r = +.22) *Word Fluency* (r = +.95) and *Digit Span* (r = +.65). In the case of *Digit Span*, however, this figure masks a decrease in s.d. between ages eight and ten. For *Perceptual Speed* the relationship is also inconsistent. Nevertheless, for *Word Fluency* the positive relationship between age and score dispersion is strong.

To summarise, score distribution decreases with age for the majority of tests in the battery. When the standard deviations of all ten tests are averaged to give a generalised picture of sample dispersion, the decrease with age is very systematic, with the correlation (-.99) approaching unity [see fig. 5.5 k].

These findings offer compelling support for the possibility that the apparent decrease in g-strength with age is an artifact of progressive range restriction.

SCORE DISPERSION AND STRENGTH OF SPEARMAN'S G.

The relationship between strength of g and score distribution (s.d.), for each test, is summarised in table 5.14. The relationship is positive in eight out of ten cases, the exceptions being Digit Span (r = -.42) and Word Fluency (r = -.96).

Table 5.14
Relationship between Standard Deviation and Strength of Spearman's g, by Two Estimates. Based on correlation matrices from five samples aged 8-12.

TEST	S.D. & Average correlation	S.D. & Percentage of variance accounted for by the first principal component.
Raven's Matrices	r = + .921 **	r = + .966 ***
Verbal Meaning	r = + .662	r = + .762
Number Facility	r = + .952 **	r = + .875 *
Spatial Relations	r = + .824 *	r = + .784
Figure Grouping	r = + .670	r = + .791
Word Grouping	r = + .725	r = + .842 *
Perceptual Speed	r = + .032	r = + .003
Word Fluency	r = - .959 ***	r = - .936 **
Digit Span	r = - .424	r = - .590
Figures of Speech	r = + .283	r = + .369
Average of SDs	r = + .884 *	r = + .943 **

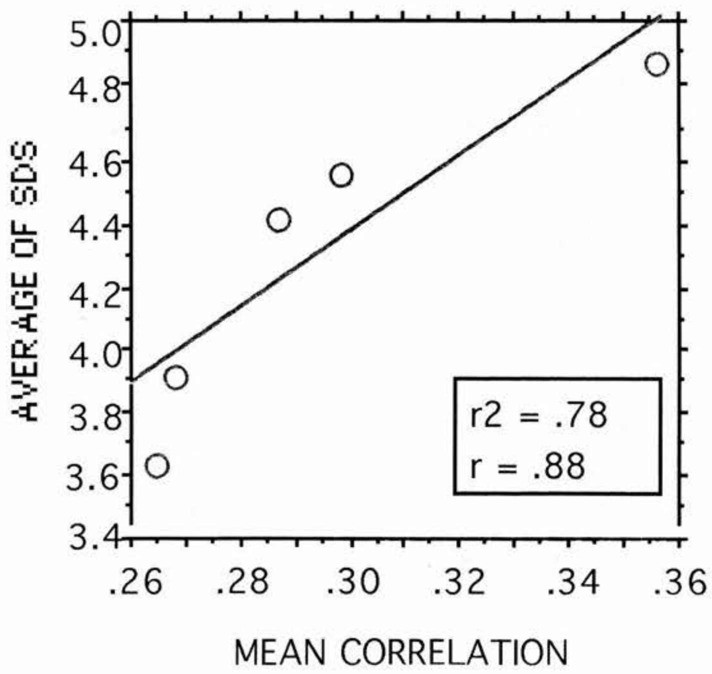
Df = 3. Critical values of r using a two-tailed test = .805 ($p<0.10$), .878 ($p>0.05$), .959 ($p<0.01$), .991 ($p<.001$)

The standard deviations of all ten tests can be averaged to give a generalised picture of sample heterogeneity. Average s.d. and g-strength (by two estimates) are plotted against each other in figures 5.6 (a) and (b). The relationship between the variables is clear (for mean correlation and average s.d., r= +.88 and for the percentage of variance explained by the first p.c. and average s.d., r = +.94), giving further support to the suggestion that declining g-strength with age is a by-product of progressive range restriction.

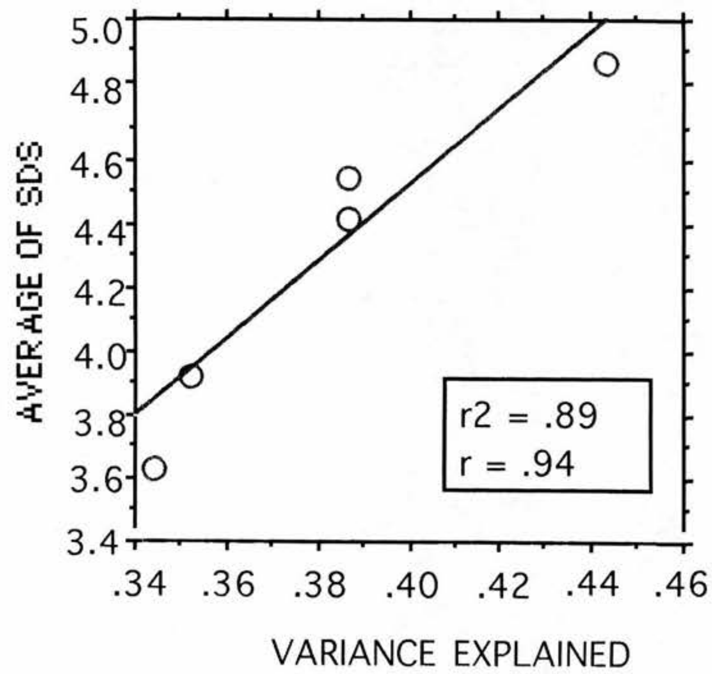
Figures 5.6 (a) & (b).

Relationship between sample heterogeneity (average of test s.d.s) and strength of Spearman's g (mean correlation, variance explained by first principal component).

a) AVERAGE OF SDS AND MEAN CORRELATION



b) AVERAGE OF SDS AND PROPORTION OF VARIANCE EXPLAINED BY THE FIRST PRINCIPAL COMPONENT



Another method of assessing the influence of progressive range restriction on g-strength across ages is to use partial correlations. Given below (table 5.15) is a matrix showing the relationship between the three variables age, mean correlation (g) and average standard deviation. To the right of this are the equivalent correlations when the effects of the third variable have been removed (table 5.16).

Table 5.15

CORRELATIONS

	AGE	MEAN R
MEAN R	-.91	
AVERAGE S.D	-.99	.88

Table 5.16

PARTIALS

	AGE	MEAN R
MEAN R	-.51	
AVERAGE S.D	-.95	-.23

As can be seen from the above, the relationship between average correlation and age drops from -.91 to -.51 when differences in standard deviation are partialled-out.

Potential Sources of Range Restriction

Genuine Restriction in Ability Range

The two indicators of sample heterogeneity which show the clearest negative relationship to age are Raven's Matrices s.d. and the average s.d. of all tests in the battery. These are also likely to be the most valid indicators of true-score ability range: Of all the tests in the battery the RPM is the one most frequently cited as a valid measure of general intelligence. The average s.d. of the battery is also likely to provide a good indication of true ability distribution within each sample.
As discussed in chapter 4, although the school from which subjects were drawn did not formally select pupils on the basis of intelligence, it is likely that incidental selection pressures operated at some level. It is also possible that these pressures increased over the course of schooling. It might be the case, for example, that with increasing school grade it becomes more likely that parents of pupils with lower-than-average performance levels will choose to withdraw them. New entrants, in contrast, may be more likely to be those who have achieved good marks in their previous schools. At the same time, it is conceivable that the parents of certain low-achievers will choose to keep their children at the school because of the specialist remedial tuition which they will receive. These conditions might explain the shift towards negative skewness found in the test score distributions of the oldest samples.

Effective Item Range / Ceiling Effects

Another potential explanation for progressive range restriction with age has to do with the tests themselves, particularly those containing items of increasing difficulty. As age increases, the proportion of items on such a test which *all* children will be able to answer, also increases. As a consequence, the effective number of items which have discriminatory power may be reduced. Range may therefore be curtailed, even though the score ceiling for the test has not been reached, although it is most likely in cases where there is a ceiling effect for some proportion of the sample.

A small proportion of subjects achieved the maximum score on four of the ten tests in the battery. These effects are summarized in table 5.17.

Table 5.17 SUMMARY OF CEILING EFFECTS

TEST	AGE	NUMBER OF SUBJECTS ACHIEVING MAXIMUM SCORE
Verbal Meaning	ELEVEN	2
Verbal Meaning	TWELVE	8
Spatial Relations	ELEVEN	2
Spatial Relations	TWELVE	3
Figure Grouping	NINE	1
Figure Grouping	TEN	4
Figure Grouping	ELEVEN	4
Figure Grouping	TWELVE	7
Word Grouping	EIGHT	1
Word Grouping	TEN	1
Word Grouping	ELEVEN	3
Word Grouping	TWELVE	5

It is clear from the above that the difficulty level of the items in these four tests does not extend far enough to discriminate between the very brightest pupils in the older age groups. Nevertheless, the number of subjects affected at each age is small, in proportion to the sample sizes, although for *Verbal Meaning* as many as eight percent of subjects are affected in the group of twelve year olds.

These findings are somewhat surprising, given that the tests were selected so as to have the widest possible discriminatory power. It is likely however, given the nature of the school, that there are a number of highly gifted pupils at each age whose scores are higher than those of the top scorers in the standardization group. While the tests are suitable for most subjects, they may be too easy for these gifted children, especially those at the highest chronological ages.

The possibility that decreases in test variances might, to an extent, be caused by curtailment in item difficulty, is supported by the observation that the three tests for which variance *increases* with age (Perceptual Speed, Digit Span, and Word Fluency) are the ones least prone to ceiling effects.*

The gradual negative skew, mentioned earlier, could also be explained in terms of spurious curtailment at the upper end of the score distribution.

It should be noted, however, that the test showing the most range restriction with age - Raven's Matrices - is clearly free of ceiling effects in the samples studied.

In summary, it seems likely that the gradual restriction of range with increasing age is primarily a result of genuine differences in ability distribution. It may also be partially due to curtailed item difficulty. Whatever the cause, the consequences are clear - that reduced correlations with age may be wholly or partly a product of selection effects. If the true extent of ability differentiation is to be revealed it is essential that this source of bias be eliminated. This procedure is the focus of the next section.

*The Perceptual Speed test contains a greater number of items than can be completed in the time allowed. For Digit Span the number sequences were increased in size beyond the capabilities of all subjects. The Word Fluency test is open-ended, with the number of potential responses constrained only by the amount of time available. Word Fluency is the test for which the increase in s.d. with age is most pronounced.

CORRECTING FOR RANGE RESTRICTION EFFECTS

STATISTICAL CONSIDERATIONS

In order to ascertain whether the apparent differentiation of abilities with increasing age reflects a genuine divergence of 'true' scores, the contaminating influence of differential test standard deviation on correlations must be removed.

There are several potential ways to approach this problem, each of which requires detailed consideration before a decision can be made as to the most appropriate and valid method. In the discussions which follow, these are split into three broad categories:

a) Selective creation of s.d.-matched samples

b) Normalisation of score distributions to produce equal means and s.d.s across tests and ages.

c) Algebraic adjustment of correlations to correct for univariate and/or multivariate selection effects.

a) CREATION OF IDEAL SAMPLES

The most obvious way to eliminate range-restriction effects, would be to selectively 'create' samples, at each age, which have identical standard deviations on all tests. The simplest method of doing this would be to compare the s.d.s of individual tests in the five samples and, for each test, to treat the sample with the smallest standard deviation as a baseline. A sub-sample from the other four age groups could then be selected in such a way as to produce score distributions with equal standard deviations for each specific test.

The main obstacle to the use of this procedure, in the current experiment, has to do with the number of suitable subjects: The creation of 'ideal' samples would involve the removal of a number of cases from each age-group. Where tests' distributions *decrease* with age, this would involve a greater reduction in numbers in younger age groups. In contrast, tests for which s.d. *increases* with age will require removal of more cases in the older age groups. Since the tests in the current battery show both increases and decreases in distribution with age, the creation of s.d.-matched samples would involve reducing sample sizes to such an extent as to render the data unusable.

This can be illustrated by considering the case of Raven's Matrices.
For the RPM the standard deviations of the eight and twelve year-old samples are 8.9 and 4.7, respectively. By gradually reducing the range of scores sampled it is possible to select a sub-group of eight year olds whose RPM s.d. (4.8) is very close to that of the twelve year olds (4.7), but whose mean score (34.4) remains similar that obtained for the whole sample of eight year olds (34.8). In order to do this, however, it is necessary to reduce the *range* of sampled scores to as few as 16, from an original range of 42. This is even smaller than the range of 26 for all twelve year-olds.[1] More importantly, the process reduces the *size* of the

sample of eight year olds to 68, from the original 111. It is clear that if the same process is applied in the case of the other variables for which s.d. decreases with age, this sample size will be reduced further.

When selecting subsamples on variables for which standard deviation *increases* with age, this reduction in numbers will be compounded since, in these cases, the compression will have most effect on the samples at the highest ages. In the case of Word Fluency, for example, the sample with the smallest s.d. (3.7) is eight-year olds. If a subsample of twelve-year olds is selected to match that s.d. the number of subjects only drops from 100 to 93, however when the range-restriction applied to eight year olds for Raven's Matrices is also taken into account, the number of subjects at age twelve reduces to a mere six.
It has already been noted that, for eight year olds the selection of a subsample equal to twelve year olds in RPM s.d. reduces the N to 68 from 111. When the range restriction for Word Fluency is also applied the sample size drops to 35.

This approach would only be useful where the samples involved are very large and where the subsamples, derived after standard deviations have been equalised, number at least 100. Deary et al. (1996) have recently used a similar method to select subsamples of children differing in age and IQ level but with equal s.d.s on a variety of subtests. Subjects were drawn from a sample of 10,500 subjects who had taken the eight subtests of the Differential Aptitude test. For each test, four idealised subgroups were specified, relating to combinations of high vs low IQ and older vs younger age. Subjects were selected using a specially designed computer program which initially selects cases meeting the specified characteristics and then makes successive adjustments to each sample (i.e. eliminates cases)

[1] The range at age twelve could, of course, be reduced and the process repeated until both range and s.d. are equal for each of the two tests at all ages. This would reduce N even further.

until all four groups have near-identical standard deviations. This process left around 300 subjects in each of the four groups (min. N =70 for "Young/Low IQ" group selected on the basis of Numerical Ability). Importantly, this figure relates to the approximate number of subjects remaining after the selection algorithm was applied with *one* test designated as the selector. It is likely that the number of selected subjects would have been substantially fewer if the criterion had been such that s.d.s had to be equal across groups for all of the individual tests, simultaneously (as would be the case in truly ideal samples). To circumvent the problem of excessively reduced sample sizes, Deary et al. selected their four groups eight times, in each case nominating one of the eight tests of the DAT as the 'selector'. For each selection, g-strength (mean correlation & percentage variance of the first principal component) was calculated for the remaining seven tests. Overall g-strength for the whole battery was estimated by averaging these eight values. It is difficult to evaluate this final estimate, however, since it is calculated on the basis of scores derived from eight *different* sets of subjects.

b) NORMALIZATION OF SCORE DISTRIBUTIONS

Another straightforward way to adjust the test score distributions to a common standard deviation at all ages, is via the process of *normalisation*. This procedure is often used to transform narrow, skewed, or otherwise uneven distributions of data points in such a way that they fit a normal curve. The normalisation process involves converting all raw scores to percentiles and then transforming these percentiles into Z-scores using the table of areas under the normal curve.*

In the applied setting, normalisation has the advantage of 'spreading-out' the scores of individuals who may be fairly homogenous for ability. This might be important, for example, in the selection of senior managers from an already highly selected pool of middle managers (Beech & Harding, 1990). The most commonly cited advantage of the procedure, however, is that it allows for the use of parametric statistics, which assume normality.

In the current context, it is easy to see how this method could be used to transform each set of test scores to a common distribution having equal means and variances at all ages.

Caution should be exercised when interpreting the statistical products of normalised data, however, since the process can distort the true nature of the data. In the current context, for example, normalisation would result in real-score intervals of the same magnitude being treated as if they were of different magnitude in different groups. For example, in the

*A variety of *algebraic* operations can also be used to increase the normality of data. The transformation used largely depends on the type of non-normality in the distribution. Positive skew, for example, can be reduced by taking the square root of each value or, in more serious cases, the logarithm. While easier to compute than the above method, none of these procedures is as effective in producing a normal distribution. More importantly, with respect to the current problem, they do not produce standardised distributions with a common mean and standard deviation across ages.

distribution for *Raven's Matrices* in twelve year olds, the 25th and 50th percentiles relate to scores of 45 and 48 respectively, a difference of *three* points. For eight year olds, however, the 25th and 50th percentiles relate to scores of 28 and 36 respectively, a difference of *eight* score points. Following normalisation, these intervals would be treated as equal. (-.67 to 0 standard deviations, or IQ 90-100 on a deviation scale with 15 points for each full s.d.) Since the RPM contains only sixty items this 'discounted' 5 points represents over 8% of the scale.

The potential distortion of true-scores constitutes the most common criticism of normalisation as a tool for use in comparative studies. This would imply that the process is inappropriate to the present analysis problem. It is arguable, however, that the normalisation may offer more advantages than disadvantages in the current context:

As already noted, one of the features of the process is that it can equalise the distributions of samples which differ in ability range. This would, apparently, allow for conclusions about age differences in g-strength to be made with the assumption of equal variances across groups. To the extent that apparent equality of variances hides differences in real-score variances, however, these conclusions must be tentative.

Although any uncertainty surrounding interpretation of results is clearly undesirable, normalisation may be a less important source of distortion than is commonly implied. Although the process equalises the scale of measurement and, to an extent, the width of that scale, it preserves the *relative* scores of subjects within each age. By spreading-out scores at the top of a distribution, it also greatly reduces the influence of curtailed item difficulty in cases where no clear ceiling effect is evident but where effective test length is reduced. This is useful to the extent that test length is related to test reliability (and consequently validity).*

Rust and Golombok (1989) argue that the claim that normalisation distorts the true nature of the data is based on a misunderstanding. They point-out that criterion referenced tests are, in reality, rarely at the interval level of measurement and that norm-referenced tests only have meaning in relation to the scores of other individuals. Because of this, the original data are not themselves 'true' and the costs of transforming them are far outweighed by the benefits. Of these, the most important is the strong advantage of having data which conforms to the assumptions of parametric statistics.

Rust and Golombok's argument that further distortion of test data is justified because they are already distorted is, perhaps, open to question. Nevertheless, their observations about the real meaning of test score intervals and the advantages of normal data for parametric analysis are very important. With respect to the differentiation hypothesis, the latter has strong advantages for the interpretation of correlational and factor analytic data.

It is clear from the above discussion that the disadvantages of normalisation may have been overestimated. Although it is not an ideal procedure for investigating the age-differentiation hypothesis it offers certain important advantages over other methods and is worthwhile considering as an option.

*Normalisation also converts scores on tests which differ in length to a scale of equal length.

Dunlap, Chen and Greer (1994) have also recently described the how normalisation can improve test reliability where distributions are moderately to seriously skewed.

c) ALGEBRAIC ADJUSTMENT OF CORRELATIONS

The methods of correcting for range-restriction discussed thus far, involve changes to the raw data itself; either through selection of idealised subsamples or transformation of scores to fit ideal criteria. It is possible that the problems associated with these methods (over-reduced sample sizes, potential scaling distortions) might be overcome by operating instead on the obtained correlations. This would involve applying one or more correcting formulae to the correlations obtained at each age, to adjust them to the levels which would be expected had the standard deviation of each test been the same in all samples.

Although no dedicated formulae have been devised to deal with this unique problem, a number of algebraic procedures exist which might usefully be applied in the current context. All of these have been designed to deal with narrow groups which have been created on the basis of scores on one or more selection tests. Because of their restricted range, the correlations obtained between test scores within such groups, will be lower than would be the case in a representative sample of the population. The correcting formulae adjust the correlations upwards, to the level that would be expected had the standard deviation of the selected group been equal to the population s.d.

These correcting procedures are variously termed 'corrections for attentuation' (as with the correction for unreliablity), 'corrections for selection effects' or 'corrections for the effects of range restriction'. Although all of the procedures have a common goal, their specific operations depend on the type of selection which has taken place.

<u>Direct, Indirect and Incidental Selection.</u>

Reduction in the size of correlations as a result of range restriction is typically a problem in samples which have been pre-selected on the basis of score on a single test (univariate selection). An example of this might an aptitude test used to select management trainees from a diverse pool of applicants. If a later test is administered to the selected group (e.g. a test of work efficiency) the correlations between that test and scores on the selector test will be lower than if all the applicants had been assessed with both measures. This is generally referred to as a *direct* selection effect, since the correlation is lowered as a consequence of direct range restriction on the selector variable. Of course, this attenuation is partly due to *indirect* restriction of range on the correlated variable, which results from direct range restriction on the selector. In the same way, the correlation between two tests, X and Y, will be attenuated if the ability range of the sample has previously been restricted via selection on a third test, Z, which correlates with X and/or Y. This situation is more commonly referred to as *incidental* selection.

Karl Pearson (1903) recognised the influence of such range restriction on the correlation coefficient, and he developed three formulae, given on the following pages, to correct for univariate selection effects. These are described thoroughly by Thorndike (1949) and Gulliksen (1950). Versions of Pearson's formulae appear in many contemporary texts on psychometrics, although rarely are they considered in any detail. Modern texts often include only the formula for correction of direct selection effects (e.g. Cronbach, 1973; Murphy and Davidshofer, 1991); or those for Pearson's cases 2 and 3 (direct and incidental selection effects) (e.g. Beech and Harding,1990). This is partly because formulae 1 and 2 (Pearson's 2 and 1) are mathematically equivalent.*

* [See information next to formula 2 and discussion on p.154]

Another reason why corrections for range restriction are so rarely discussed in modern texts is that, while most researchers are aware of their existence, few are experienced in using them. Ree et al. (1994) point-out that between 1988 and 1992 only 4% of the validity articles in the journals *Educational and Psychological Measurement, Journal of Applied Psychology* and *Personnel Psychology* corrected for range restriction. This oversight is reinforced by the absence of such correction procedures in common statistical packages.

Perhaps the main reason why correcting procedures are so rarely used, and why knowledge about them has waned, is that they are often impossible to compute. This is because information about test standard deviations in the unrestricted sample is seldom available. When the formulae *are* used, the s.d. in the unrestricted group is usually taken as the s.d. of the test's standardisation sample, where this is given in the test manual. Of course, this is only an estimate, since the sample under consideration is unlikely to have been selected from the standardisation group.

In the *current* study the standard deviations of all tests in all age groups and in the whole sample are known. The conversion of correlations at each age to a common s.d. for each test does not, therefore, require whole population estimates.

PEARSON'S FORMULAE FOR CORRECTING CORRELATION COEFFICIENTS FOR THE EFFECTS OF DIRECT, INDIRECT, AND INCIDENTAL RANGE RESTRICTION.

1. Formula to Correct for the Effects of Direct Selection (Pearson's "Case 2")

Applies when the correlation to be corrected is between two variables X and Y, selection occurred on variable X and restricted and unrestricted variances are known only for X.

$$rXY^{*} \qquad \frac{r\,XY \times (SX^{*}/SX)}{\sqrt{[1-\ rXY^{2} + [rXY^{2} \times (SX^{*}/SX)^{2}]}}$$

Where: rXY^{*} = The correlation between X & Y corrected for the effects of range restriction

rXY = the correlation between the selector test (X) and another test (Y) obtained in the selected (range-restricted) group.

SX^{*} = The standard deviation of the selector test in the 'population' (the unselected/unrestricted group)

SX = The standard deviation of the selector test in the range-restricted group.

2. Formula to Correct for Pearson's "Case 1". (Rarely used.)

Applies when the correlation to be corrected is between two variables X and Y, selection occurred on variable X and restricted and unrestricted variances are known only for Y.

Assuming that the regression of X on Y is the same in the restricted and unrestricted samples, the outcome of this formulae will be identical to that for correction of direct selection effects.

$$rXY^{*} = \sqrt{1-(SY^{2}/SY^{2*})(1-rXY^{2})}$$

Where: rXY^{*} and rXY are as above

SY^{2*} = The variance of the scores on test Y in the 'population' (the unselected group)

SY^{2} = The variance of test Y in the selected group

Formula to Correct for the Effects of Incidental Selection (Pearson's Case 3.).

Applies when the correlation to be corrected is between two variables Y and Z, the sample has previously been selected on the basis of another variable X and restricted and unrestricted variances are known only for X

$$rZY^* = \frac{rZY + rXY \times rXZ \times U}{\sqrt{[1 + rXY^2 \times U] \times [1 + rXZ^2 \times U]}}$$

where: r_{ZY}^* = The correlation between tests Z & Y *corrected* for range restriction

$r\,ZY$ = The correlation between tests Z & Y obtained in the selected (range restricted) sample

SX^* = s.d. of scores on test X in the unrestricted group

SX = s.d. of scores on test X in the selected (range restricted) group

$U = ((SX^*/SX)^2 - 1)$

$r\,XY$ = The correlation between tests X & Y for the selected (range restricted) sample

$r\,XZ$ = the correlation between tests X & Z for the selected sample.

The above formulae are based on the usual rules of correlation, whereby the unknown s.d. of a test can be predicted from its correlation with another test for which the standard deviation is known. By the same principles, the unknown, unrestricted s.d. of a criterion test Y can be predicted from its known correlation with a selector test X in selected group and the restricted and unrestricted s.d.s of X (formula 1). The unknown, unrestricted s.d. of X can be predicted from rXY and the restricted and unrestricted s.d.s of Y (formula 2). In the formula for incidental selection effects, the unrestricted s.ds of Y & Z are derived from rXZ, rXY and restricted and unrestricted variances of X. In all cases corrected correlations are based on estimates of 'unrestricted' test s.d.s

The central determinant of correction in the formulae for direct and incidental selection is the ratio of selector test standard deviation in the range-restricted group to selector test s.d. in the unrestricted group. Y substitutes for X in formula 2 which is mathematically equivalent to the formula for direct selection.

Assumption of Univariate Selection

The assumption of univariate selection has important implications with respect to the applicability of Pearson's formulae to the current problem. Although the standard deviations of individual tests clearly differ across age groups, this is not the result of *explicit* range-restriction, neither does it stem from a *single* selection test.

Nevertheless, if it can be established that the data behave in much the same way as would be the case if direct univariate selection *had* taken place, then there may be grounds for using Pearson's formulae.

Do Age Groups Resemble Selected Sub-Samples?

As was discussed earlier, when the standard deviations of all tests are averaged to give an indication of general ability distribution, there is a clear decrease in distribution with age. This is also the case for the Raven's Matrices test. The fact that these two estimates of general intelligence show so systematic a decline in standard deviation with age implies that there is a genuine restriction of ability range in the samples. For the purposes of correction the whole sample might therefore be considered the 'population' or 'unselected' group. This has an advantage over the usual method of using the standardisation sample as the hypothetical population, in that the subgroups are drawn from the same pool of subjects.*

Since equivalence over ages is the main consideration, it might also be feasible to treat the sample of eight year olds as the 'unselected' group, since they have least restricted range of scores on these variables.

*This is, in fact, debatable. It is discussed in more detail in a later section.

Does one test behave as the key selector?

Of all the tests in the battery, Raven's Matrices is the one most often referred to in the literature as a test of general intelligence. In the present study the RPM is also the test showing the strongest negative relationship between s.d. and age, despite an apparent lack of ceiling effects. Since explicit restriction in the range of general ability within a sample, would lead to reductions in the ranges of all other tests with a loading on g, it could be argued that the RPM resembles a selection test.

In view of the above, it might seem reasonable to treat the RPM as as the selection test and the whole sample or the sample of eight year olds as the non range-restricted group. Correlations of the RPM with other variables could then be corrected using the formula for *direct* range restriction effects and the other correlations could be corrected using the formula for *incidental* range restriction.

Problems with treating Raven's Matrices as the selector.

Unfortunately, closer examination of the data reveals that this apparently convenient solution is, in fact, inappropriate and is likely to lead to *overcorrection* of the correlations. This is because the main determinant of correction in Pearson's formulae is the ratio of selector s.d. in the restricted vs unrestricted groups. The corrections adjust all correlations upwards in proportion to this ratio and since this ratio is highest in the RPM, the majority of correlations will be over-adjusted.
In addition, although standard deviations decrease with age for the majority of tests in the battery, they*increase* for three of the tests. Corrections using RPM as the selector will exaggerate this existing increase and spuriously raise the g-saturation of the matrix even further. This problem was confirmed when the initial correlation matrices were corrected for direct and incidental range-restriction using the RPM as the selector and the sample of eight year olds as the 'population'. This

resulted in overcorrection to such an extent that the relationship between age and g-strength was reversed.[1] This finding would have been similar if the whole sample had been treated as the unselected group, since the s.d. of RPM is similar in the whole sample (8.6) to the eight year old subgroup (8.9).[2]

The tendency of corrections for range-restriction to over-correct, is noted by Frearson et al. (1988). These investigators first computed correlations between psychometric and psychophysical tasks in a widely dispersed sample of 97 subjects and then in four smaller sub-groups selected on the basis of global IQ score. Using the restricted and unrestricted s.d.s of the selector criterion they then calculated estimated 'population' correlations using Pearson's procedure(s)[3]. For each of the four groups, these 'corrected' correlations, were compared to the actual correlations obtained in the unrestricted sample. In all cases, the correcting procedure had boosted the correlations above their reference values. This finding is consistent with the hypothesis that the ratio of standard deviations before and after range restriction is likely to be highest in the selector itself, since other tests are imperfectly correlated with it. Adjusting all other correlations to the level determined by this high ratio is likely to overcompensate for the effects of range restriction.

1 These calculations were, in fact, made to the correlation matrices which existed before the raw data was re-marked for the purposes of reliability estimation. This process revealed a few minor marking errors which required all calculations to be made again. The pattern of g-strength with age in the uncorrected data was very similar to that given in the present chapter and it seems plausible to assume that corrections for range restriction would result in the same reversal effect in the current (slightly modified) data. Since the correction procedure is very time consuming and ultimately flawed, the costs of re-calculating corrections were not warranted.

2 [Figures for the other age groups are: Nine 7.3, ten 7.1, eleven 5.9, twelve 4.8]

3 Actual procedure not specified

Assumption of Multivariate Selection

From the above discussion, it would seem that there are few grounds for treating individual age groups as subsamples selected on the basis of a single test. It would therefore appear inappropriate to use Pearson's correcting formulae to equalise age groups for test variances.
The problem with the current data is that the degree and type (positive/ negative) of association between age and score distribution differs between tests. It is therefore impossible to identify a single test as the key selector for all of the samples. In effect, each of the variables in each of the samples has been incidentally selected on the basis of some unknown dimension which relates age to ability range. In addition, within each age group and in the sample as a whole, all of the tests potentially exert selection effects.
If all tests are exerting selection effects, the standard deviation of each variable will be modified to a level determined by the cumulative range-restricting effects of all the other variables. An ideal correcting procedure would, therefore, have to adjust each correlation for the effects of range-restriction stemming, simultaneously, from all the tests. In order that age groups might be compared, it would also have to adjust the correlations at each age to a common standard deviation.

While no currently available formula would meet all these requirements, one existing procedure may offer a partial solution. Lawley's (1943) procedure for multivariate selection was designed to correct for the effects of range restriction stemming from several selection tests (often referred to as the 'general case').

As Gulliksen has pointed-out, *"the equations for multivariate selection in the general case [are] almost prohibitively complex"* (1950, p.158). For this reason, Lawley's multivariate correction procedure relies on matrix algebra. This is, itself, fairly daunting to the uninitiated. Lawley's correction procedure is described overleaf.

Lawley's (1943) Procedure for Correcting Correlations for the Effects of Multivariate Selection.

Suppose that a sample has been selected on the basis of scores on a battery of *p* variables for which population information is available. These subjects are then assessed with another set of tests, equal to *n - p* (total number of tests minus the selection tests). The Variance-covariance matrix for the restricted sample is:

$$V^* = \begin{matrix} V^*_{p,p} & V^*_{p,\ n-p} \\ V^*_{n-p,\ p} & V^*_{n-p,\ n-p} \end{matrix}$$

All of the above values are known.

The comparable variance-covariance matrix for the unrestricted group (the 'population') is:

$$V = \begin{matrix} V_{p,p} & V_{p,\ n-p} \\ V_{n-p,\ p} & V_{n-p,\ n-p} \end{matrix}$$

Where there is only knowledge of Vp,p (the variance & covariance of the selector tests).

The known variances and covariances in the unrestricted sample (Vp,p) plus the variances and covariances of the restricted sample (all known) can be used to provide corrected variances and covariances for all variables, using the following matrix equations:

$$V_{p,\ n-p} = V_{p,p}\ V^*_{p,p}\ V^*_{p,n-p},$$

and $$V_{n-p,\ n-p} = V^*_{n-p,\ n-p} + V^*_{n-p,p}\ V^*_{p,p}\ (V_{p,\ n-p} - V^*_{p,\ n-p})$$

which can be combined to give

$$V_{n-p,\ n-p} = V^*_{n-p,\ n-p} + V^*_{n-p,p}\ V^*_{p,p}\ (V_{p,p}\ V^*_{p,p} - i_p)\ V^*_{p,n-p}$$

where i_p is a $p \times p$ matrix with ones on the diagonal and zeros on all off-diagonal entries (the identity matrix).

The corrected variances and covariances can than be converted to correlations.

Lawley's method effectively takes into account *direct* selection of sub-groups on several variables and *incidental* selection of the remaining non-selector variables. With certain assumptions (see below) it also takes into account *indirect* selection effects within subgroups. In these respects, Lawley's procedure is similar to Pearson's univariate method, but with the advantage of being able to deal simultaneously with several selector tests whose potential range-restricting effects may vary.

Unfortunately, Lawley's method also retains many of the problems of Pearson's univariate method in terms of its usefulness to the present investigation.

As with Pearson's method, Lawley's procedure is designed for cases of explicit selection. For the method to be used in the correction of correlations obtained in different age groups a decision must be made as to which tests to treat as "the selection battery". In the extreme case, nine of the ten tests could be considered selectors and one the criterion test, although this begs the converse question of which test to consider not a selector. The other important point about using Lawley's procedures is that in the current sample all values in the variance-covariance matrix for the 'unrestricted' group are known (assuming this to be the entire sample). This begs the question of what values we are trying to predict with the matrix equations.

The difficulties with univariate and multivariate correction procedures, with respect to their applicability to the current problem, hinge on the two key mathematical assumptions underlying them. These are the assumption of linearity and the assumption of equality of variances.

Two Key Assumptions Underlying Algebraic Corrections for Range Restriction.

Assumption of Linearity.

This is the assumption that, for any pair of correlated variables, X and Y, there exists a constant linear relationship. The slope of the 'true' regression line of X on Y will therefore be the same irrespective of the portion of either variable which is examined. In other words, it is assumed that the true relationship between X and Y is the same in a range-restricted sample as in the population. (The size of the correlation coefficient will, nevertheless, be reduced as a result of range restriction.) If this assumption holds, then the correlation between X and Y in the 'population' (i.e. the corrected rXY) can be estimated from information about the restricted and unrestricted variances of X and the correlation between X and Y in the range-restricted group.

Assumption of Equality of Error Variance

The assumption of a constant linear relationship between X and Y only leads to the prediction of 'population' or *corrected* rXY where one of two conditions applies: Either (a) both variables are entirely composed of true-score variance, or (b) the proportion of variance in either variable which is due to measurement error, is the same in the unselected and selected groups. Since (a) is extremely unlikely, correcting procedures must assume that for any test, the proportion of variance which is due to *variability of measurement error* is the same across the entire distribution. It should also be assumed that the error variances of X and Y are uncorrelated, such that the *covariance* of X & Y is entirely due to true scores.

Problems with assumming linearity across age groups.

The assumption of linearity lies at the heart of all procedures for estimating range-corrected correlations. It also constitutes the main obstacle to the use of these procedures for removing the effects of differential range-restriction from the correlations obtained within each age group. This obstacle operates at two levels:

Theoretical:

The assumption of linearity is, in fact, antithetical to the differentiation hypothesis, which assumes that the true relationship between ability variables *differs* across the range of possible scores (Deary et al., 1996). Specifically, the differentiation hypothesis would predict that the degree of linearity, in a regression of one ability variable on another, reduces as age increases. Furthermore, this is hypothesised to stem from changes in true score variances, rather than differences in error variances.

Practical:

Because linearity cannot be assumed, the technique of using information about test variances and covariances in one sample to predict them in another sample, is brought into question.

In the usual situation where corrections for range-restriction are used (i.e. in cases of explicit selection), there are grounds for assuming that the direct selection of subgroups on certain variables will have predictable indirect and incidental effects on others. Specifically, range restriction on one or more variables will have range-restricting effects on other, correlated variables. This is not the case in the current data, however. Although average test variance decreases with age, suggesting selection, the variances of several tests increase or are largely unchanged with age, despite remaining correlated with the tests most resembling 'selectors'.

This lack of linearity could be taken to imply that the age-subgroups were not drawn from the same population.[1] This interpretation accords with the differentiation hypothesis in so far as as the latter predicts qualitative differences in the strength of g across ages. Before this can be accepted, however, it is important to check whether the lack of a systematic relationship between age and variance across variables (such as might be expected in the case of explicit selection) is genuine, or whether it is due to error variance.

Can equality of error variances be assumed across age groups?

It is a widely held assumption in psychometrics, that error variance will be equal across samples, irrespective of their heterogeneity.[2] The term *error variance* does not relate to the particular sources of error, but rather to the variability of their effects on the sample. Consider, for example, two groups of twenty subjects who are tested in rooms of different size. In a small room all subjects may hear the test instructions clearly, whereas in a very large room, subjects at the back may not hear the instructions as well as those at the front. The proportion of total variance which is due to variability in error will be greater in the latter group. Equality of variances is a plausible assumption because there is no reason to suppose that a subgroup, selected on the basis of test score(s), will be subjected to more variable testing conditions than a sample of unselected subjects.

[1] (or, technically, that the distributions of test scores within each age group were not drawn from the same superordinate bivariate distribution).

[2] Within traditional psychometrics, the idea that error variance is not affected by changes in group heterogeneity is moot. On the one hand, much of the work of early psychometricians such as Kelly (1921) and Otis (1922) is based on the assumption that error variance will be curtailed to the same extent as overall variance, following selection. Others, such as Gulliksen (1950) and Green (1951), are open to the possibility that the standard error of measurement might be different after range-restriction. The latter also argue that error variance may be the same in sub-groups whose test scores differ in mean, variance and reliability.

Sample Dispersion and Test Reliability

An added complication in regard to algebraic corrections for range restriction, is the question of whether to correct the *obtained* correlations or the correlations which have previously been *corrected* for test unreliability.

There are arguments for and against either approach. On the one hand, it could be argued that corrections for unreliability have removed one source of contamination from the matrices at each age and should be used in conjunction with corrections for range restriction, which remove another, more important, source. On the other hand, since corrections for unreliability have little effect on the age-g relationship it is questionable whether the procedure is necessary. Furthermore, many psychometricians (e.g. Kline, 1989) urge caution in the use of such procedures, arguing that they may actually distort the true data. To reduce the likelihood of such artificial distortions it is, therefore, important to minimise the number of transformations made to the data.

An alternative interpretation of the reliability estimates complicates matters even further. As noted by Gulliksen (1950) *reliability estimates are affected by sample heterogeneity.* It is possible, therefore, that corrections for unreliability have, to an extent, already corrected for range-restriction. According to Gulliksen, the following formula can be used to estimate the amount of change in reliability that would be expected from a change in group variance, assuming that the entire change is due to a difference in true score variance:

$$RXX = 1 - \frac{1 - sx^2}{SX^2}(1 - rxx)$$

Where: RXX is the 'corrected' reliability estimate
rxx is the obtained reliability estimate
sx^2 is the variance of the test in the range-restricted sample
SX^2 is the variance of the test in the unrestricted sample.

As with corrections for range-restriction, the usefulness of the above formula is constrained by its assumption of linearity and equality of error variances. (Hence its reliance on the ratio of variances in the hypothetically selected and unselected groups.) In this regard it is not clear how the 'corrected' reliability estimates and the 'corrected' correlations based on these 'corrected' reliability estimates, could be interpreted.

Nevertheless, it is important to check the degree to which sample heterogeneity may have affected the reliability estimates obtained at each age. In particular, it must be ascertained whether the relationship between age and reliability-corrected estimates of g-strength, changes when the reliability estimates have themselves been corrected for the effects of range restriction:

Using Gulliksen's correcting formula, range-corrected estimates of test reliability were calculated for each age group. These are compared to uncorrected reliability estimates in table 5.18.

Table 5.18 Estimates of test reliability across age groups, corrected for range restriction. (Uncorrected reliability estimates are shown in brackets.)

	EIGHT	NINE	TEN	ELEVEN	TWELVE
Raven's Matrices	.93(.92)	.89(.92)	.90(.93)	.79(.90)	.82(.94)
Verbal Meaning	.89(.94)	.88(.92)	.92(.96)	.89(.96)	.84(.96)
Number Facility	.93(.94)	.91(.95)	.91(.97)	.92(.97)	.96(.98)
Spatial Relations	.78(.80)	.77(.83)	.67(.80)	.80(.85)	.74(.87)
Figure Grouping	.60(.62)	.69(.73)	.81(.82)	.64(.78)	.48(.82)
Word Grouping	.66(.68)	.69(.76)	.76(.77)	.65(.86)	.57(.87)
Perceptual Speed	.91(.95)	.91(.95)	.93(.96)	.94(.96)	.93(.96)
Word Fluency	.95(.98)	.96(.97)	.92(.94)	.95(.96)	.94(.96)
AVERAGE	.85(.83)	.86(.83)	.88(.84)	.90(.82)	.91(.80)

As can be seen from table 5.18, when the range-corrected figures for all tests are averaged, the decline in reliability with age, found in the original estimates, becomes an increase. This accords with the common observation of reduced measurement error with increasing age (e.g. Anastasi, 1979). The differences between the two estimates of reliability also increase with age, in line with an increase in the ratio of restricted to unrestricted variances. Despite these features, the differences between the range-corrected and uncorrected reliability estimates are small.

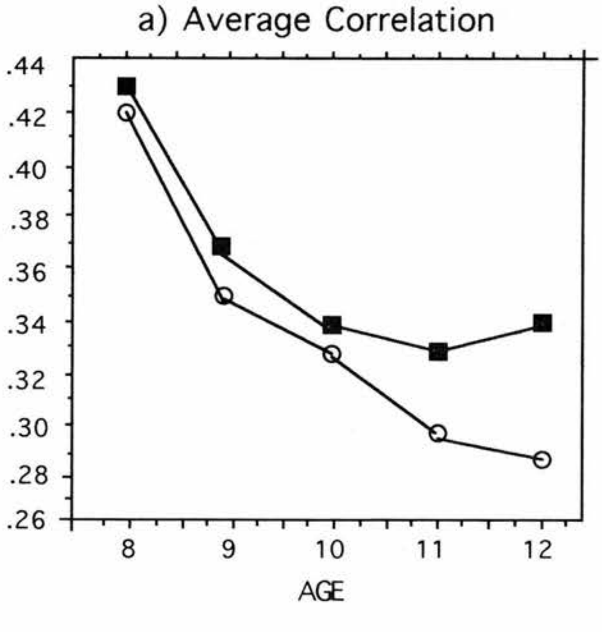
Reliability-corrected estimates of g-strength, within the five age groups, are shown in table 5.19 and figures 5.7 a and b. In each case, results derived from range-corrected reliability estimates are compared to those previously reported for correlations corrected according to the original reliability estimates (p. 129, fig. 5.4 a, tb. 5.10).

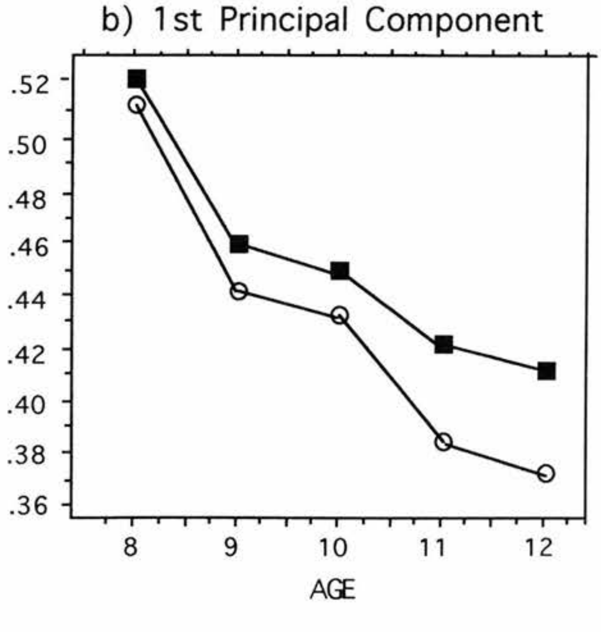
Table 5.19 Strength of g, following correction for unreliability, using range-corrected versus original estimates of test reliability.

AGE	MEAN r Corrected using range-adjusted reliabilities	MEAN r Corrected using Original reliabilities	VARIANCE 1ST PC Corrected using Range-corrected reliabilities	VARIANCE 1ST PC Corrected using original reliabilities
EIGHT	.42	.43	.51	.52
NINE	.35	.37	.44	.45
TEN	.33	.34	.43	.44
ELEVEN	.30	.33	.38	.42
TWELVE	.29	.34	.37	.41

Figures 5.7 (a) and (b). Age and Strength of g by two estimates, following correction for unreliability, where reliability estimates have, or have not, been corrected for range restriction.

■ = Corrected using original reliability estimates
○ = Corrected using range-adjusted reliability estimates





As can be seen from these figures and tables, using range-corrected reliability estimates to correct correlations for the attenuating effects of test unreliability, does not alter the general pattern of reduced g-strength with age. It does result in a slightly stronger relationship, however, since range-correction raises reliability estimates as age increases (and range decreases). The obtained correlations are, therefore, less likely to be raised by the correction for attenuation in older samples.*

It has already been noted that correcting for test unreliability has little effect on the relationship between age and g. The same procedure, using range-corrected reliability estimates, has even smaller effects. In view of the minor effects of such procedures, and bearing in mind the problems associated with multiple corrections, it is concluded that no advantage is offered by using reliability-corrected correlations to compute algebraic corrections for range restriction.

*For g estimated as average correlation, the relationship with age changes from -.91 to -.94. For g estimated as the first principal component, the relationship to age changes from -.94 to -.96.

RESULTS FOR STUDY 1, USING NORMALIZED DATA.

As discussed earlier, the normalisation process involves the conversion of raw scores to percentile scores, followed by transformation of the latter to Z-scores, based on the statistical table of areas under the perfect normal curve. By fitting the existing data to a normal curve, the effects of outlyers and of skewness are eliminated and narrowly distributed scores can be spread-out over a wider range of values. The Z-transformation also converts all distributions to a common scale with the same mean and standard deviation. In contrast to the Z-transformation of *raw* scores, however - where the 'identical' standard deviations are merely nominal and the shape and true variance of the distribution remain unchanged - the Z scores derived from *normalised* data relate to samples which have been fitted to the same normal curve.

Normalisation of Data

Scores on all tests were normalised and standardised within each age group, using the method described above, as performed by the 'minitab' statistics package. To bring all values to the positive, normalised standard scores were converted to IQs distributed around a mean of 100, with a standard deviation of 15. (This is achieved by multiplying the Z-scores by 15 and adding 100). All resultant IQ distributions matched these figures when rounded upwards to the nearest whole number. The exception to this was Digit Span, for which the standard deviation of the normalised IQs is close to 13.5 in all five age groups. This smaller s.d. simply reflects the lesser potential variability of the raw data.

Relationship between raw and transformed scores.

Before considering age differences in estimates of g-strength, derived from normalised data, it is important to ascertain the extent of agreement between normalised and obtained scores. The lower the concordance between these scores, the more likely it is that the transformation process has distorted the true nature of the data. This would invalidate age comparisons based on estimates of g obtained from normalised scores. Close agreement between normalised and obtained scores, however, would imply that the process has effectively preserved the true-score distribution of each variable, whilst equalising spread and eliminating the effects of skewness and outlyers.

Correlations between normalised scores and raw-scores are shown below:

Table 5.20 Concordance Between Normalised and Raw Scores.

	EIGHT	NINE	TEN	ELEVEN	TWELVE
Raven's Matrices	.988	.970	.955	.964	.993
Verbal Meaning	.993	.998	.967	.972	.917
Number Facility	.983	.990	.993	.995	.996
Spatial Relations	.988	.996	.994	.995	.995
Figure Grouping	.996	.959	.982	.982	.984
Word Grouping	.995	.981	.957	.991	.986
Perceptual Speed	.989	.985	.992	.993	.993
Word Fluency	.996	.993	.993	.999	.993
Digit Span	.992	.998	.999	1.000	1.000
Figures of Speech	.997	.998	.994	.997	.990

As can be seen from table 5.20, the extent of agreement between raw scores and normalised scores approaches unity in all cases. The concordance levels far exceed the reliabilities of all the relevant tests, indicating that very little information has been lost as a result of the transformation process. The normalised/raw data concordance level is lowest for *Verbal Meaning* at age twelve (.92). In view of the reported ceiling effect for this test at this age, however, it could be argued that the normalised distribution is a better reflection of 'true' score dispersion.

Age Differences In the Strength of Spearman's g, Using Normalised Data.

Intercorrelations between normalised scores on the ten subtests, within each of the five age groups, are shown in appendix 5.3.

Age differences in the strength of Spearman's g, estimated as average correlation and proportion of total variance accounted for by the first unrotated principal component, are shown table 5.21 The relationship between age and g is illustrated graphically in figure 5.8.

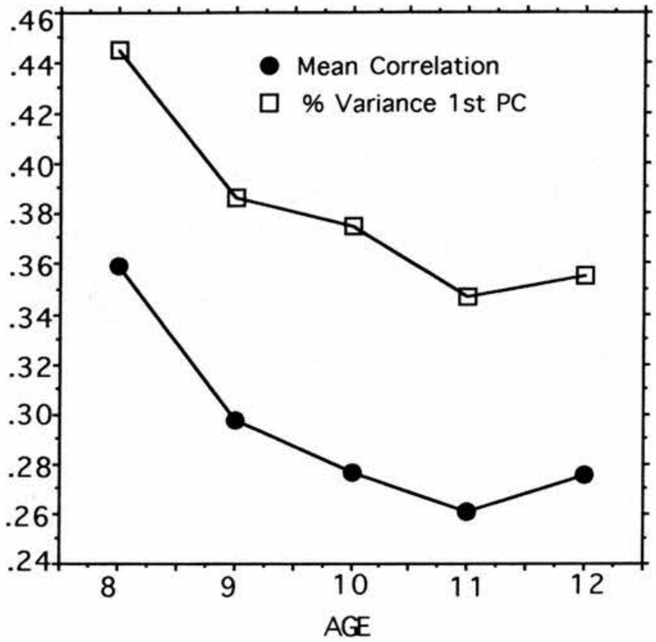
Table 5.21 AGE AND STRENGTH OF G, USING NORMALISED DATA.

AGE	Mean Correlation (excluding diagonals)	Percentage of Variance Explained by the First Principal Component
8	.359	.445
9	.298	.387
10	.277	.375
11	.261	.347
12	.276	.355

As can be seen from table 5.21 and figure 5.8 the relationship between g-strength and age remains negative, despite normalisation of the data sets.

The relationship between age and g-strength using normalised data, is compared to that observed using raw (non-normalised) data, in figures 5.9 (a) and (b). From these it can be seen that normalisation has had little effect on the size of the average correlation coefficient and the first principal component. It does, however, change the decrease in g-strength between the ages of eleven and twelve, to a slight increase.

FIGURE 5.8 AGE AND STRENGTH OF SPEARMAN'S G, BY TWO ESTIMATES, USING NORMALISED DATA



FIGURES 5.9 (a) & (b). COMPARISON OF THE RELATIONSHIP BETWEEN AGE AND G, USING RAW VERSUS NORMALISED DATA.

Fig 5.9 (a). Age and Mean Correlation, Using Raw versus Normalised Data

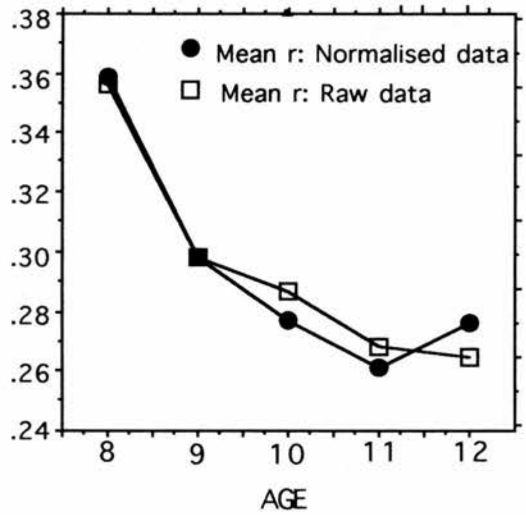
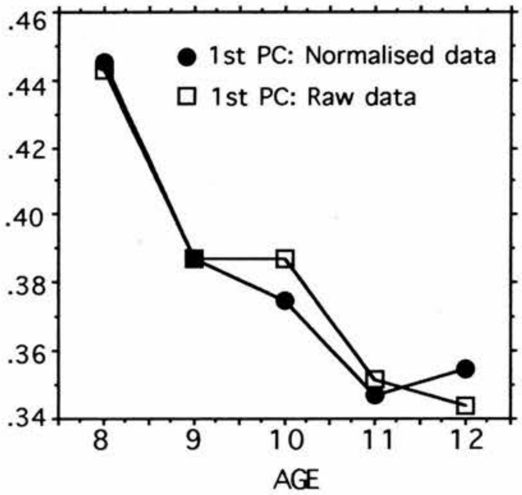


Fig 5.9 (b) Age and Percentage of Variance Accounted for by the First Principal Component



That normalisation should have little effect on g-strength in the youngest samples was to be expected, given the existing normality of the raw data. The failure of normalisation to reduce the extent of decrease in g-strength with age, however, is surprising in view of the apparent effect of group heterogeneity on correlation size, suggested by the partial correlations reported earlier. Given the potential of normalisation to distort the true nature of the data, however, it is important to consider whether there is an alternative explanation for these findings.

In this regard, it is must be determined whether differences in sample heterogeneity continue to play a part in the decline of g-strength with age, despite normalisation. As already noted, the Z transformation, in itself, only nominally changes distributions to the same mean and standard deviation. With normalised data, however, Z-scores were mapped from percentile scores, using a table of areas under the perfect normal curve. Because the distribution of obtained scores has been spread-out in this way, there are grounds for treating the normalised data sets across ages as having equal distributions. This can only be assumed, however, where there is sufficient differentiation between scores within each distribution, such that scores can be split into percentiles. If this is the case, IQs derived from normalised standard scores will range between 62 and 138 in samples of this size. These figures relate to Z-scores of -2.53 and +2.53, respectively, relating to the lower and upper levels within which an estimated 99.9 percent of obtained scores fall. In cases where there is poor differentiation between raw scores, however, the <u>potential</u> spread of transformed scores will be constrained, since a larger percentage of subjects will be represented within each score interval, making it difficult to split the sample into percentile scores. This is likely where raw scores are distributed within a narrow range covering only a small number of values, such as where a test has only a small number of items, or where the distribution of test scores is leptokurtic, skewed or curtailed by ceiling effects. The latter will result in a higher percentage of subjects achieving the top score, than would be expected in a normal distribution. This makes it impossible to calculate values within the upper percentiles. Thus, while normalisation will be expected to greatly reduce sample differences in score spread, it cannot be counted-on to eliminate these differences entirely. The raw scores of the older subjects in the current sample are more narrowly distributed, <u>and</u> more prone to curtailment at the upper levels, than those of the younger subjects. The possibility must therefore be considered, that normalisation may be less effective in spreading-out scores, as age increases. One way to check this is to compare the ranges of normalised scores in the five age groups. These are shown in table 5.22.

<u>Table 5.22 Range of Normalised IQ Scores Across Ages and Tests</u>

	EIGHT	N INE	TEN	ELEVEN	TWELVE
Raven's Matrices	71 (67-138)	71(62-133)	70(63-132)	75(62-138)	72(63-134)
Verbal Meaning	76 (62-138)	73(62-135)	75(63-137)	72(62-134)	64(63-126)
Number Facility	71 (62-132)	76(62-138)	75 (63-137)	75(62-138)	72(63-134)
Spatial Relations	76 (62-138)	71(62-133)	70(63-132)	67(68-134)	70(63-132)
Figure Grouping	73 (62-135)	76(62-138)	68(63-130)	68(62-131)	64(63-127)
Word Grouping	76 (62-138)	68(62-130)	72(66-137)	75(68-132)	67(63-129)
Perceptual Speed	76 (62-138)	71(62-133)	72 (66-137)	75(62-138)	75(63-137)
Word Fluency	73 (65-138)	76(62-138)	75(63-137)	75(62-138)	70(63-132)
Digit Span	76 (62-138)	61(70-131)	64(63-126)	53(77-129)	64(73-137)
Figures of Speech	68 (67-135)	71(67-138)	70(68-137)	72(66-138)	72(66-137)
MEAN RANGE*	74	71	71	70	69

As can be seen from the above, the ranges of normalised IQ scores vary between tests, with a maximum range of 76 (min: 62, max:138). As expected, Digit Span scores convert to the smallest range of IQs, since there are only a small number of 'items' for this test.

Shown in the bottom row of table 5.22 is the average range of normalised IQ scores at each age. From this it can be seen that the average range of scores decreases from 74 at age eight to 69 at age twelve. These figures suggest a decline in true score variance with age, despite apparent equality of variance. For this reason the results using normalised data must be interpreted with caution.

AGE DIFFERENCES IN THE STRENGTH OF SPEARMAN'S G, FOLLOWING ALGEBRAIC CORRECTION FOR THE EFFECTS OF RANGE RESTRICTION.

The earlier discussion of Pearson and Lawley's formulae for correction of univariate and multivariate selection effects, highlighted a number of problems concerning their use in compensating for age differences in sample dispersion. These can be summarised as follows:

Assumption of linearity:

The correcting formulae operate on the assumption that the degree of relationship between any pair of variables is the same across the entire range of each. Observation of the correlations in the whole sample and the sub-samples at each age, however, reveals that this is not the case. It is also assumed that restriction in one variable will always lead to a predictable degree of restriction in the range of correlated variables. When the variances of the individual tests across age groups are compared, however, it is clear that the extent of change in variance with age, differs between tests.

Assumption of a formal selector.

The key determinant of correction, using Pearson's formulae, is the ratio of restricted to unrestricted variances in one or more selection tests. No specific selector has been used in the current study, however, neither does the data behave as if it had.

Since it is not possible to identify a specific selector or group of selectors it is essential that each correlation be treated individually. In this regard, the formulae for correction of incidental and multivariate selection effects are unsuitable, since they assume that attenuation in the correlation between two variables, X and Y, is a consequence of selection on a separate variable (Z) or variables (Z_1, Z_2...).

As linearity cannot be assumed, the formula for correction of direct selection effects will also be unsuitable, in its usual form, since it requires that one variable be nominated the 'selector'. Given the nature of the data in the current study, the best estimate of unrestricted correlation will obtained by taking into account the potential attenuating effects of range <u>restriction in both variables</u>.

To meet the above requirements, the following method was used:

For each pair of variables, two estimates of 'unrestricted' correlation were calculated, using Pearson's formula for the correction of direct selection effects. Each of these estimates was calculated treating one of the variables in the pair as the 'selector' and the whole sample as the 'population'. For each pair of variables, the final range-corrected correlation is taken as the average of the two estimates.

Tables 5.23 a. to e. contain both estimates of range-corrected correlation, with the coefficients in each column corrected with the assumption of the common variable as the 'selector'. If the data conformed exactly to the principles of linearity, these matrices would be symmetrical. As can be seen, however, this is not the case.

TABLES 5.23 a) to e).

CORRELATIONS AT EACH AGE, CORRECTED FOR THE EFFECTS OF ATTENUATION DUE TO RANGE RESTRICTION.

PART 1:
MATRICES CONTAINING TWO ESTIMATES OF UNRESTRICTED CORRELATION, EACH TREATING ONE VARIABLE IN THE CORRELATED PAIR AS THE SELECTOR.

For each column, the common variable is treated as the selector (X) in calculating the 'corrected' intercorrelations. This yields two estimates for each original coefficient: For example, with eight year olds (below) the correlation between Raven's Matrices and Verbal Meaning is adjusted to .443 when RPM is treated as the direct selector (column 1) and to .338 when VM is nominated the selector (column 2).

a. EIGHT YEAR OLDS - ASYMMETRICAL MATRIX.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.443	.562	.537	.451	.468	.483	.441	.112	.249
Verbal Meaning	.338	*	.647	.415	.248	.557	.492	.642	.078	.315
Number Facility	.458	.654	*	.590	.432	.570	.642	.719	.168	.464
Spatial Relations	.508	.504	.679	*	.422	.48	.604	.484	.064	.388
Figure Grouping	.425	.312	.500	.424	*	.344	.367	.303	-.083	.268
Word Grouping	.443	.653	.660	.482	.344	*	.522	.59	.212	.395
Perceptual Speed	.382	.503	.646	.522	.302	.441	*	.537	.194	.385
Word Fluency	.328	.632	.711	.389	.235	.485	.516	*	.228	.417
Digit Span	.093	.090	.163	.058	-.074	.190	.215	.266	*	.146
Figures of Speech	.227	.383	.526	.381	.261	.385	.451	.507	.159	*

b. NINE YEAR OLDS- ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.537	.578	.570	.438	.436	.355	.197	.269	.218
Verbal Meaning	.512	*	.583	.420	.414	.504	.393	.191	.174	.250
Number Facility	.501	.531	*	.441	.490	.637	.666	.259	.200	.260
Spatial Relations	.570	.443	.514	*	.410	.437	.342	.249	.077	.251
Figure Grouping	.467	.466	.598	.438	*	.543	.387	.037	.01	.235
Word Grouping	.446	.540	.72	.447	.523	*	.572	.393	.033	.374
Perceptual Speed	.301	.355	.672	.290	.307	.492	*	.278	.194	.273
Word Fluency	.183	.19	.292	.232	.032	.361	.307	*	.130	.385
Digit Span	.265	.183	.238	.076	.009	.031	.228	.138	*	.167
Figures of Speech	.259	.314	.367	.298	.259	.426	.379	.474	.202	*

c. TEN YEAR OLDS - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.709	.653	.578	.561	.56	.39	.187	.137	.11
Verbal Meaning	.67	*	.576	.497	.442	.527	.235	.41	.166	.092
Number Facility	.534	.499	*	.495	.502	.426	.622	.433	.126	.272
Spatial Relations	.553	.513	.589	*	.493	.389	.511	.279	.121	.122
Figure Grouping	.626	.546	.685	.583	*	.525	.424	.245	.061	.017
Word Grouping	.626	.634	.606	.472	.525	*	.44	.127	.096	.035
Perceptual Speed	.365	.242	.708	.507	.343	.357	*	.352	.113	.149
Word Fluency	.19	.454	.555	.302	.212	.109	.382	*	.076	.312
Digit Span	.133	.179	.166	.125	.05	.078	.119	.072	*	.163
Figures of Speech	.132	.124	.422	.156	.017	.035	.193	.363	.202	*

d. ELEVEN - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.512	.644	.496	.523	.637	.419	.233	.412	.17
Verbal Meaning	.44	*	.456	.291	.337	.445	.351	.34	.302	.122
Number Facility	.611	.496	*	.26	.292	.486	.547	.418	.287	.104
Spatial Relations	.58	.419	.344	*	.426	.506	.474	.12	.254	.306
Figure Grouping	.568	.439	.35	.391	*	.647	.312	.074	.247	.167
Word Grouping	.605	.486	.487	.397	.571	*	.498	.198	.206	.197
Perceptual Speed	.469	.465	.634	.444	.319	.583	*	.366	.233	.294
Word Fluency	.281	.473	.522	.117	.081	.259	.385	*	-.111	.325
Digit Span	.523	.463	.404	.274	.293	.296	.271	-.123	*	.153
Figures of Speech	.202	.177	.135	.295	.178	.253	.306	.32	.136	*

e. TWELVE - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.629	.584	.471	.533	.466	.479	.339	.198	.353
Verbal Meaning	.581	*	.164	.552	.219	.447	.184	.384	.153	.348
Number Facility	.638	.213	*	.263	.487	.255	.623	.374	.253	.319
Spatial Relations	.572	.7	.295	*	.462	.211	.363	.322	.034	.231
Figure Grouping	.566	.267	.466	.399	*	.555	.286	.263	.192	.171
Word Grouping	.47	.498	.226	.166	.528	*	.331	.372	.154	.176
Perceptual Speed	.602	.281	.691	.381	.354	.432	*	.486	.262	.215
Word Fluency	.498	.601	.488	.384	.371	.533	.54	*	.235	.427
Digit Span	.331	.293	.367	.045	.298	.259	.323	.255	*	.220
Figures of Speech	.539	.583	.445	.296	.262	.288	.262	.45	.216	*

Discrepancies among the estimates based on alternative 'selectors', are summarised, in the form of averages, in table 5.24. Shown, for each sample, are two estimates of the average correlation of each test with the other nine. The first of these is calculated from 'corrected' correlations, where the named variable is treated as the 'selector' (X). Shown in brackets is the average inter-correlation when the named variable is treated as the criterion (Y) and each of the other variables is treated as the selector.

Table 5.24 AVERAGE CORRELATION OF EACH TEST WITH THE OTHER NINE, WHEN THE NAMED TEST IS TREATED AS THE SELECTOR (X), VERSUS THE CRITERION (Y).

	EIGHT		NINE		TEN		ELEVEN		TWELVE		Row Average	
	X	(Y)	X	(Y)	X	(Y)	X	(Y)	X	(Y)	X	(Y)
Raven's Matrices	.42	(.36)	.40	(.39)	.43	(.43)	.45	(.48)	.45	(.53)	.43	(.44)
Verbal Meaning	.43	(.45)	.39	(.39)	.41	(.43)	.35	(.43)	.34	(.45)	.38	(.43)
Number Facility	.53	(.56)	.46	(.49)	.46	(.53)	.39	(.44)	.37	(.43)	.44	(.49)
Spatial Relations	.44	(.44)	.36	(.37)	.39	(.42)	.35	(.36)	.32	(.36)	.37	(.39)
Figure Grouping	.31	(.30)	.33	(.34)	.36	(.40)	.34	(.35)	.35	(.39)	.34	(.36)
Word Grouping	.46	(.45)	.44	(.44)	.35	(.38)	.42	(.44)	.33	(.38)	.40	(.42)
Perceptual Speed	.47	(.44)	.38	(.37)	.36	(.36)	.39	(.43)	.36	(.43)	.39	(.41)
Word Fluency	.49	(.45)	.24	(.25)	.27	(.29)	.22	(.26)	.36	(.46)	.32	(.34)
Digit Span	.12	(.13)	.14	(.15)	.12	(.13)	.22	(.28)	.19	(.27)	.16	(.19)
Figures of Speech	.34	(.36)	.27	(.33)	.14	(.18)	.20	(.22)	.27	(.37)	.24	(.29)
Column Average	.40	(.39)	.34	(.35)	.33	(.36)	.33	(.37)	.33	(.41)	.35	(.38)
Average Difference	.01		.01		.03		.04		.08		.03	

As already noted, for the individual pairs of range-corrected correlations, there are clear (although small) differences between the estimates, depending on which of the two contributory variables is treated as the selector in their calculation. It will be noted from table 5.24, however, that *on average* within and across samples, the differences between these two estimates are very much smaller. This indicates that there is some linearity in the data, and offers support for the use of Pearson's formulae in correcting the data.

Interestingly, the differences between the averages of these estimates increases with age, suggesting a decrease in linearity, as predicted by the differention hypothesis.

Final estimates of range-corrected correlations (i.e. the averages of the two previous estimates) are shown in appendix 5.4 (part 2).

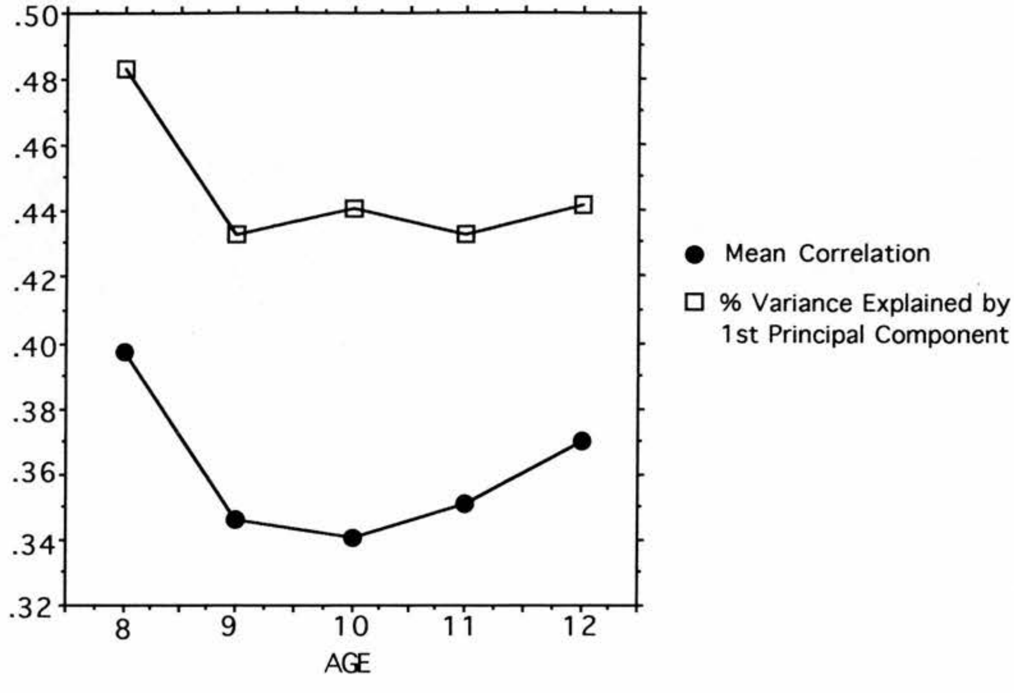
As with previous versions of the data, the strength of Spearman's g, at each age, was estimated as the average of the matrix (excluding the diagonals) and the proportion of variance explained by the first unrotated factor. Age differences g-strength, by these two estimates, are shown in table 5.25 and figure 5.10.

Table 5.25 <u>Age and Strength of Spearman's g, After Algebraic Correction for Range Restriction.</u>

AGE	Mean correlation (excluding diagonals)	% Variance explained 1st principal component
EIGHT	.397	.483
NINE	.346	.433
TEN	.341	.441
ELEVEN	.351	.433
TWELVE	.370	.442

Patterns of g-strength across age-groups, <u>before and after</u> correction for range restriction, are compared in figures 5.11 (a) and (b)

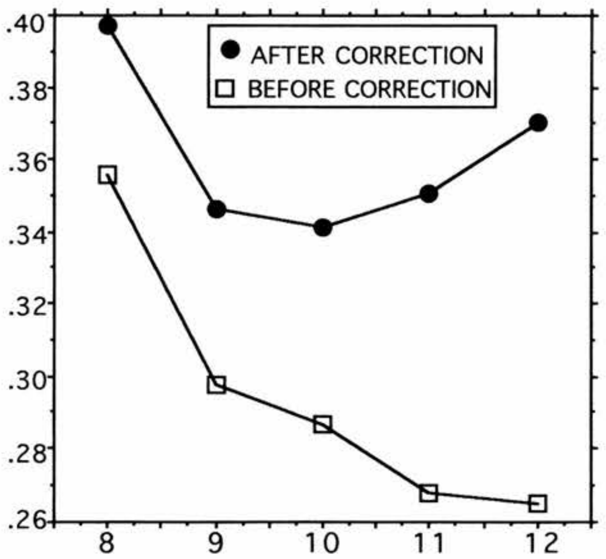
FIGURE 5.10. AGE AND STRENGTH OF G BY TWO ESTIMATES, AFTER ALGEBRAIC CORRECTION FOR THE EFFECTS OF RANGE RESTRICTION

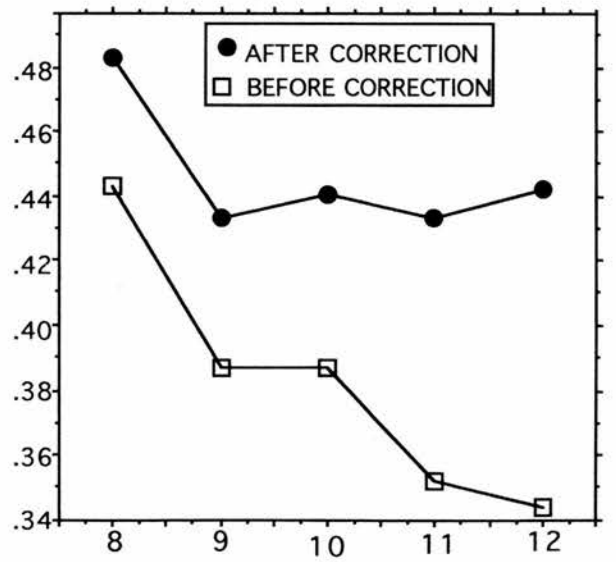


FIGURES 5.11 a & b. AGE AND STRENGTH OF G, BEFORE AND AFTER ALGEBRAIC CORRECTION FOR THE EFFECTS OF RANGE RESTRICTION

Fig. 5.11 a. Age and Mean Correlation

Fig. 5.11 b. Age and Percentage of Variance Explained by the First Principal Component.





As can be seen from table 5.25 and figures 5.10 and 5.11 a & b, the negative relationship between age and g-strength, found in the original data, is greatly diminished after correcting for the effects of range restriction. For variance explained by the first principal component, the relationship falls from -.94 to -.62. The latter correlation is mainly due to the high relative value of the first eigenvalue in the sample of eight year olds. There is, however, no consistent trend in g-strength, by this estimate, between the ages of nine and twelve.
For the average correlation, the relationship with age falls from -.91 to -.33 after correcting for range-restriction. As with the previous estimate, the relationship between mean r and age is largely explained by the relatively high value for the eight year old sample. Interestingly, however, there appears to be a curvilinear relationship between age and mean correlation, with a decrease between the ages of eight and ten, followed by an increase to age twelve.

The large decrease in the correlation between age ang g-strength, following correction for range restriction, is in line with earlier predictions, based on partial correlations between age and g, with the influence of sample heterogeneity removed. Although a negative relationship remains, it has already been noted that this is explained by the relatively high g-strength of the sample of eight year olds. When only considering the samples ranging between ages nine and twelve, the relationship of g to age is, in fact positive (age & mean r: +.84, age & first PC: +.50). It should be noted, however, that these correlations are based on small differences in actual values.

Given the lack of consistency in the relationship between age and g, it could be concluded that the original differentiation found in the data, is explained almost entirely by differences in sample heterogeneity.
An alternative explanation, however, focuses on problems with the correcting formula. As discussed earlier, a number of writers on

psychometrics (e.g. Kline, 1993; Ree et al., 1995; Frearson, et al.,1988) have expressed reservations about the use of such procedures to correct for attentuation. For example, Ree et al. (1994) note that application of Pearson's formulae may actually result in reversal of the *sign* of correlations and Frearson et al. (1988) have documented the tendency of Pearson's formulae to *overcompensate* for range-restriction. Although the signs of the correlations in the range-corrected matrices have not changed as a result of the correcting procedures, the 'corrected' pattern of g-strength across age-groups could, conceivably, be due to the type of overcompensation noted by Frearson et al. As discussed earlier, the key determinant of correction in this formula is the ratio of variances in the 'selected' versus 'unselected' samples. If corrections overcorrect, then the degree of overcorrection will be proportional to the degree of range restriction in the sample. Since the ratio of restricted to unrestricted variances increases with age, it might be expected that the overcorrections would have proportionally greater effects, the older the samples. This could explain the curvilinear relationship betwen age and mean correlation noted above.

Summary of Results for Study 1.

Results based on the unaltered raw data obtained for the five samples, indicate a gradual decline in the strength of g with increasing age, as predicted by the differentiation hypothesis.

On average, test reliability and score variance also decline as age increases, suggesting that the reduction in g-strength may be artifactual.

Age differences in reliability are small, however, and correcting the correlational data for the effects of differential reliability does not alter the general direction of the results. This continues to be the case when reliability estimates are corrected for the effects of range restriction, although the decrease in reliability with age, becomes an increase.

Normalisation of the data within each age group, eliminates skew and minimises age differences in score spread. This has little effect on the negative relationship between age and g. It is difficult to evaluate these results, however, since normalisation may distort the 'true' scores, hence the apparent equalisation of distributions may be illusory.

Algebraic correction for range-restriction effects, using Pearson's formula, substantially reduces the negative relationship between age and g-strength. The slight relationship which remains is accounted for by the relatively high values of the g estimates in the youngest sample. These results would imply that the progressive weakening of positive manifold with increasing age is an artifact of range-restriction effects. This conclusion cannot, however, be made with total confidence, since Pearson's formula may itself produce artifact by overcompensating for range restriction. The U-shaped relationship between age and average correlation, following Pearson's 'corrections', supports this possibility.

CHAPTER 6

STUDY 2: EXAMINING THE STRENGTH OF SPEARMAN'S G AT DIFFERENT ABILITY LEVELS.

This chapter will examine the psychometric data for evidence of ability differentiation at higher IQ levels. The aim of the first analysis is to identify general trends in g-strength within samples of equal spread drawn from the sections above and below the mean population IQ. In the second analysis, the sample will be divided into smaller groups comprising a larger number of narrower IQ bands.

GENERAL STATISTICAL APPROACH

REMOVAL OF AGE VARIANCE

Before addressing the relationship between g-strength and ability level, it is first necessary to remove age variance from the data. This was achieved by standardising scores on each of the psychometric tests, within age groups. Rather than transforming raw scores directly to z-scores, they were first converted to percentile scores and then to z-scores based on those expected in an ideal normal distribution. Normalised standard scores were then converted to IQ scores with a mean of 100 and a standard deviation of 15 in each of the five age groups. As discussed in the previous chapter, normalised and non-normalised scores correlate very closely within these samples, but the former are more suitable for parametric analysis, since skewness and outliers have been eliminated. Most importantly, in the present context, this method of standardisation produces distributions of approximately equal shape and spread across age-groups and tests.

A cautionary note.

As described in chapter 4, when mean IQ levels are estimated on the basis of published conversion tables (Raven's Matrices 1979; Thurstone's PMA, 1962) they increase slightly with age (RPM 110-115; PMA 110-113). It could be argued, therefore, that the mean IQ within each sample, nominally set at 100, should be adjusted to reflect these differences. Normative data is not available for all of the tests, however, so the appropriate degree of adjustment would have to be estimated in those cases. In view of the imprecise nature of such adjustments and the small difference between age-groups in the average of the five PMA tests (max. 3 points), it was thought best to treat the data as if the means (100) relate to equivalent ability levels in all of the subsamples.

Features of the IQ Distribution

The distribution of normalised IQ scores for nine of the ten tests in the battery falls within the range 62 to 138 (61.93 - 138.07), equivalent to Z-scores of -2.53 to +2.53, or the lower and upper levels within which 99.9 percent of obtained scores fall. The IQ range for the *Figures of Speech* test (65 to 138) is slightly curtailed at the lower end, reducing the range to approximately 73 points. Although the Z-score to IQ transformation will ideally produce a standard deviation of 15 IQ points, insufficient delineation between percentile scores at the lower and upper tails results in slightly smaller values, with the s.d.s of IQ for most of the tests falling between 14.5 and 14.7.

Choice of selection procedure

The primary aim of these analyses, is to compare the strength of Spearman's g in samples which are equal with respect to ability range, but which differ in average IQ. In order to accomplish this, it is first necessary to identify a criterion of selection. One obvious way to split the sample, would be to average the ten IQs obtained by each subject and then to divide the entire sample according to specified ranges of global IQ. As noted by Detterman and Daniel (1989) and Deary et al. (1996) however, the use of average scores is likely to lead to an unacceptably high level of attenuation in the correlations between sub-tests. According to Detterman and Daniel, the best method of selecting sub-samples for this type of study, is to use an independent selection test. Since the object of their investigation was the standardisation sample for the Weschsler batteries, however, no such test data was available to them. Instead, they chose to split the sample into different IQ ranges using either one of two subtests (Vocabulary or Information) as the criterion of selection. The authors report that selection by different subtests only produces minor differences in final estimates of g. This conclusion, however, is based on only two of the 11 or 12 subtests in the WAIS-R and WISC-R. In contrast, Deary et al. note that inter-subtest correlations and estimates of g in high and low IQ groups will differ depending on the subtest which is used for selection. Of particular influence, is the test's correlation with the general factor. The higher the g-loading of the variable used for selection, the more likely it is that its restriction will have simultaneous effects on other ability variables, reducing the size of their inter-correlations. In view of this, it can be argued that a more reliable estimate of g-strength at each ability level, will be obtained by *averaging the estimates obtained in sub-samples selected by each of the variables.* This requires the computation of a separate correlation matrix for each selected subgroup (N matrices = N groups x N variables).

Should the selector be included or excluded when estimating g?

In Detterman and Daniel's study, scores on all of the battery subtests, *including the direct selector,* were correlated in order to calculate estimates of g-strength. In contrast, Deary et al. *do not include* the selector variable in their matrices. Although Detterman and Daniel report that inclusion of the selector in the analysis makes little difference to the results, there is an important statistical reason for excluding it: The distribution of the selector variable is specifically curtailed by the selection procedure, hence its normality is likely to be seriously affected, making it unsuitable for parametric analysis. Although the distributions of the other variables are also likely to be be narrowed as a consequence of incidental selection, the degree of this restriction will be far lower, since they are imperfectly correlated with the direct selector. The normality of the non-selector variables is therefore less likely to be compromised by the selection process. (Evidence for this is supplied in a later section.)

For the above reasons, the following approach will be adopted in the coming analyses:

In short, the sample will be split into sub-groups according to IQ on each of the variables. Correlation matrices will then be calculated for each sub-group, excluding the criterion test. Thus, of the ten matrices computed for every sample, each test will be included in nine out ten analyses and act as the selection criterion in one out of ten. Finally, the inter-subtest correlations obtained in the ten subsamples for each IQ range will be averaged, to give an overall picture of g-strength.

STUDY 2 A: STRENGTH OF SPEARMAN'S G IN SAMPLES BELOW AND ABOVE MEAN IQ.

The aim of this analysis is to examine the strength of Spearman's g within sub-samples of equal spread, drawn from the sections above and below the mean sample IQ. This will allow for the identification of general differences between two large and widely dispersed samples with mean IQs which differ from each other but which are within the typical range.

SAMPLE SELECTION.

In order to create range-matched groups above and below mean IQ, samples were selected so as to include subjects scoring within the IQ ranges 62 - 99.5 ('LOW') and 100.5 - 138 ('HIGH'). Subjects scoring at the mean of 100 were eliminated, in order to delineate the samples. The IQ range within each sample (37.5 points) is therefore equivalent to around 2.5 full-sample standard deviations.

Following the procedure described above, a separate selection of 'low' and 'high' IQ samples was made using each of the tests in the battery as the criterion. This selection method produces ten samples of 'low' scorers and ten samples of 'high' scorers.

The mean, range and standard deviation of IQ scores in the 'high' and 'low' groups selected by each of the variables, are shown in table 6.1. Also shown are the sample sizes.

Table 6.1 Range and standard deviation of IQ scores in samples below and above the mean, and in the whole sample.

SELECTION CRITERION	MEAN IQ Overall=100		RANGE			STANDARD DEVIATION			SAMPLE SIZE Total N=527	
	Low	High	Low	High	All	Low	High	All	Low	High
Raven's Matrices	87.8	112.2	36.6	37.4	76.0	8.7	8.6	14.7	256	255
Verbal Meaning	87.7	112.0	37.2	37.0	76.0	8.7	8.6	14.6	253	258
Number Facility	87.9	111.8	37.4	37.0	76.1	8.7	8.6	14.7	257	262
Spatial Relations	87.7	111.5	36.6	37.4	76.0	8.6	8.7	14.6	252	269
Figure Grouping	88.5	112.6	37.0	36.2	76.1	8.8	8.0	14.4	268	241
Word Grouping	88.1	112.2	37.2	37.3	76.0	8.8	8.2	14.5	258	256
Perceptual Speed	87.9	112.2	36.0	37.0	76.0	8.7	8.7	14.7	256	264
Word Fluency	87.6	111.5	37.1	37.0	76.1	8.6	8.8	14.7	250	269
Digit Span	89.0	112.3	35.5	35.5	75.9	7.7	7.8	13.5	257	232
Figures of Speech	88.2	111.9	34.0	36.4	72.5	8.7	8.5	14.6	262	256
Average	88.0	112.0	36.5	36.8	75.7	8.6	8.5	14.5	257	255

As can be seen from table 6.1, the average IQs of the samples below and above the mean are 88 and 112, respectively, a difference of 24 points. The range and standard deviation within the selected 'high' and 'low' IQ samples are approximately equal, although there are minor variations, depending on which test is used as the selector. On average, the range of scores within each sample represents approximately 48% of the population range and the standard deviation around 58% of the population s.d. These figures are close to those reported by Deary et al. (1996), whose 'Dull and 'Bright' sample means were 90 and 110, respectively and whose distribution is described as being "about half the population standard deviation" (Data from the most recent Standardisation of the Differential Aptitude test.)

In summary, the selection procedure has apparently* produced two groups with widely differing mean IQs but approximately equal score ranges.

*Discussed later

Normality of selector vs non selector variables.

As already noted, each of the ten subtests in the battery was, in turn, used as the criterion of selection, with samples split at the centre of the distribution. As a consequence, the distributions of the selector variables are clearly skewed. Skewness coefficients for the selectors in each of the samples range from a miminum of +/- .77 to a maximum of +/-.94. In all cases, these values indicate a significant departure from normality (5% confidence level, one-tailed). In contrast, the distributions of the remaining nine variables remain normal, irrespective of the selection criterion, with none of the coefficients of skewness exceeding +/- .29 (n.s.).

Skewness and kurtosis coefficients for all variables in the 'low' and 'high' groups, selected by each criterion, are given in appendix 6.1 Histograms given in appendix 6.2 illustrate the different distributions of the selector and non selector variables, using the example of sub-groups selected by Raven's Matrices IQ. It can be seen that while the distributions of Raven's Matrices IQ in the 'low' and 'high' groups resemble two halves of a bell, the bell curves of the other tests remain intact. This pattern in evident for all the other selection tests, indicating that non-selector variables are appropriate for parametric analysis. As already stated, the direct selectors were excluded from analyses performed for each sub-group.

RESULTS FOR STUDY 2A:

Twenty matrices of inter-test correlations were calculated, relating to the 'high' and 'low' ability samples selected according to IQ on each of the variables. In each case, the relevant selector test was excluded from the matrix. The individual matrices are tabulated in appendix 6.3.

In all groups, correlations are smaller in size than would have been expected in an unrestricted sample. As the selected range of IQ is equal in 'high' and 'low' ability groups, however, this does not represent a problem for their comparison.

Summarised in table 6.2, are the average correlations of each variable with the other eight in the 'high' and 'low' IQ groups selected by the tenth variable. As already stated, the variable which has acted as the selector is not included in the matrix. Diagonals are excluded before calculating averages.

Table 6.2

Average correlation of each variable with the other eight in range-matched 'low' and 'high' ability sub-samples selected by IQ on each of the ten tests.

SELECTOR:	RPM		VM		NF		SR		FG		WG		PS		WF		DS		FS	
SUBGROUP:	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI
Variable																				
Raven's Matrices	--	--	.27	.20	.29	.28	.22	.23	.19	.17	.26	.26	.28	.26	.21	.23	.13	.09	.18	.20
Verbal Meaning	.32	.27	--	--	.32	.33	.28	.24	.24	.23	.28	.25	.30	.28	.21	.21	.10	.11	.22	.20
Number Facility	.28	.24	.26	.25	-	--	.24	.26	.22	.19	.24	.26	.24	.18	.21	.14	.10	.09	.15	.20
Spatial Relations	.27	.30	.28	.25	.32	.34	--	--	.21	.24	.31	.28	.25	.29	.23	.24	.13	.14	.16	.24
Figure Grouping	.29	.31	.30	.29	.34	.34	.25	.30	--	--	.29	.29	.31	.29	.26	.27	.15	.18	.17	.27
Word Grouping	.26	.29	.27	.19	.29	.30	.27	.25	.21	.20	--	-	.25	.25	.23	.20	.12	.11	.19	.22
Perceptual Speed	.31	.33	.30	.29	.33	.28	.28	.26	.27	.24	.29	.28	--	-	.25	.16	.13	.12	.21	.19
Word Fluency	.34	.36	.26	.30	.33	.35	.30	.31	.28	.30	.30	.33	.31	.27	--	--	.15	.13	.11	.24
Digit Span	.40	.30	.35	.31	.41	.34	.37	.32	.31	.27	.37	.33	.34	.32	.30	.26	--	-	.21	.24
Figures of Speech	.35	.36	.31	.31	.35	.37	.29	.32	.29	.28	.34	.32	.33	.29	.21	.22	.13	.11	--	--

Although the above table summarises the information contained in the twenty correlation matrices, this information cannot be usefully interpreted without further condensation.

In line with the procedures adopted in the analysis of age-related differentiation, the amount of common variance in each selected sub-group was first estimated by averaging the correlations in the relevant matrix, excluding the diagonals. The second estimate of g to be calculated is the proportion of variance accounted for by the first unrotated principal component. These estimates of g-strength, for each of the twenty subgroups, are shown in table 6.3.

<u>Table 6.3 Strength of Spearman's g by two estimates, in 'high and 'low' ability groups selected by IQ score on each of the ten tests. (Selector excluded)</u>

SELECTION TEST	MEAN CORRELATION		VARIANCE EXPLAINED BY FACTOR 1	
	Low IQ	High IQ	Low IQ	High IQ
Raven's Matrices	.22	.21	.32	.31
Verbal Meaning	.25	.24	.35	.33
Number Facility	.22	.21	.31	.30
Spatial Relations	.24	.26	.34	.35
Figure Grouping	.26	.28	.35	.37
Word Grouping	.23	.22	.33	.32
Perceptual Speed	.26	.24	.35	.34
Word Fluency	.27	.29	.36	.38
Digit Span	.34	.30	.42	.38
Figures of Speech	.29	.29	.38	.38

It is apparent from table 6.3 that estimates of g-strength are similar in the 'low' and 'high' IQ groups, irrespective of the criterion of selection. Between groups, the difference in average correlation and variance explained by the first principal component, does not exceed .04 Estimates of g in the sample below mean IQ exceed those in the sample above mean IQ in only six out of ten cases.

There is more variability among estimates of g between samples selected using the ten different tests, although these differences are also small.

This variability is to be expected, since a slightly different group of tests was included in the computation of each pair of matrices, and since the g-loading of each selector is different. Table 6.4 shows the loading of each test on the first unrotated factor, for each subsample. From this it can be seen that the tests with the most consistently high loadings on the general factor are Raven's Matrices and Number Facility. These also represent the selection tests which yield matrices with the smallest amount of common variance. In contrast, the g-loadings of Digit Span tend to be the lowest, and it is where Digit Span is used to split the sample, that the greatest amount of common variance remains in the matrices.

Table 6.4. Loading of each variable on the first unrotated factor in 'low' and 'high' IQ sub-samples selected by IQ on each of the ten tests.

SELECTOR:	RPM		VM		NF		SR		FG		WG		PS		WF		DS		FS	
SUBGROUP:	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI	LO	HI
TEST																				
Raven's Matrices	--	--	.71	.64	.69	.64	.64	.67	.65	.66	.63	.70	.68	.75	.72	.74	.75	.63	.72	.73
Verbal Meaning	.65	.51	--	--	65	63	.67	.55	.66	.61	.65	.47	.67	.66	.58	.63	.66	.63	.64	.65
Number Facility	.69	.69	.71	.75	--	--	.73	.75	.74	.71	.67	.72	.70	.64	.71	.72	.75	.69	.71	.75
Spatial Relations	.55	.58	.65	.59	.64	.68	--	--	.56	.64	.65	.62	.62	.64	.65	.66	.69	.65	.62	.68
Figure Grouping	.49	.47	.57	.58	.59	.55	.52	.56	--	--	.53	.54	.62	.60	.64	.64	.61	.57	.62	.61
Word Grouping	.63	.66	.64	.60	.60	.67	.71	.64	.63	.62	--	-	.65	.68	.68	.68	.69	.67	.70	.67
Perceptual Speed	.67	.65	.66	.66	.58	.48	.59	.65	.67	.62	.60	.62	--	--	.67	.57	.64	.66	.67	.60
Word Fluency	.53	.59	.48	.50	.51	.39	.55	.55	.58	.59	.56	.48	.55	.37	--	--	.56	.53	.46	.47
Digit Span	.32	.23	.24	.28	.26	.26	.30	.33	.35	.39	.29	.27	.30	.29	.34	.28	--	-	.30	.24
Figures of Speech	.45	.50	.49	.45	.38	.50	.38	.53	.39	.56	.46	.52	.46	.43	.25	.51	.41	.49	--	--

The significance of the difference between correlations in the 'high' and 'low' IQ samples, was tested using the statistical methods reported by Detterman and Daniel (1989). Correlations within each matrix were first converted to a normal distribution using Fisher's r to Z transformation. Fisher's z test was then applied to the difference between the corresponding correlations in the matrices obtained by the 'high' and 'low' ability samples. Values of z exceeding plus or minus 1.96 indicate that the correlations are significantly different at the 0.5 level, using a two-tailed test. In their study, Detterman and Daniel next counted the number of statistically significant differences in each pair of matrices and compared these to the number (5%) expected by chance, using the chi squared test.

Shown in table 6.5 are the number of correlations, out of a total of 72 (one triangular portion of a 9 by 9 matrix excluding the diagonals), which are significantly different in the high and low ability groups (Fisher's z).

Table 6.5 Number of significant differences between corresponding correlations in the matrices obtained by 'high' vs 'low' IQ samples, selected according to each test.

Selection Test	N. Significant differences (z) between High and Low IQ Samples.
Raven's Matrices:	0
Verbal Meaning:	2
Number Facility:	2
Spatial Relations:	2
Figure Grouping:	1
Word Grouping:	1
Perceptual Speed:	1
Word Fluency:	3
Digit Span:	3
Figures of Speech:	1

Out of 72 pairs of correlations, only a maximum of three show differences above the chance level, according to Fisher's z test. There is little point in conducting further chi square tests, since up to 3.6 pairs of correlations (5%) are expected to be different by chance.

The results, thus reported, offer no support for the differentiation hypothesis. As mentioned in the introductory section, however, a better estimate of g- strength is likely to be obtained by averaging the corresponding correlations obtained in the ten 'high' IQ and ten 'low' IQ samples, to produce two final matrices. These are combined in table 6.6, with the 'high' IQ sample represented in the top right-hand section (bold type) and the 'low' IQ sample represented in the lower left-hand section (plain text).

Table 6.6 Averages of corresponding correlations in all 'high' IQ **(bold typeface)** and 'low' IQ (normal typeface) subgroups.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	**.32**	**.41**	**.45**	**.39**	**.37**	**.30**	**.14**	**.16**	**.19**
Verbal Meaning	.41	*	**.28**	**.28**	**.25**	**.36**	**.20**	**.29**	**.18**	**.22**
Number Facility	.41	.34	*	**.29**	**.30**	**.41**	**.46**	**.33**	**.15**	**.30**
Spatial Relations	.40	.37	.32	*	**.37**	**.30**	**.33**	**.18**	**.06**	**.23**
Figure Grouping	.40	.25	.34	.34	*	**.35**	**.20**	**.08**	**.01**	**.15**
Word Grouping	.37	.41	.35	.28	.37	*	**.30**	**.21**	**.12**	**.17**
Perceptual Speed	.27	.23	.48	.32	.24	.35	*	**.25**	**.16**	**.21**
Word Fluency	.20	.30	.34	.18	.12	.25	.32	*	**.07**	**.35**
Digit Span	.21	.12	.18	.08	.06	.11	.16	.08	*	**.17**
Figures of Speech	.15	.16	.20	.20	.10	.18	.21	.29	.14	*

The estimates of g-strength derived from these matrices are shown in table 6.7.

Table 6.7 Strength of Spearman's g by two estimates, in averaged 'high' and 'low' IQ samples.

	'LOW' IQ	'HIGH' IQ
Mean inter-subtest correlation, excluding diagonals.	.25	.26
Percent of total variance explained by the first principal component.	.34	.34

It is clear, from the above, that there is no difference in the strength of g between the 'high' and 'low' IQ samples. When the two matrices of averaged correlations were compared using Fisher's z, no significant differences between corresponding correlations were found.

TEST RELIABILITY IN THE 'HIGH' AND 'LOW' IQ SAMPLES.

Although the 'high' and 'low' IQ samples have been matched for range, group differences in test reliability remain a potential source of error in the correlational data.

Reliability estimates for each test in the twenty 'high' and 'low' IQ subgroups, are shown in appendix 6.4. (Calculations follow the procedures detailed in the last chapter.) No reliability estimate is provided for a test where it is acting as the selection criterion for the sample, since it will not have been included in the relevant correlational analysis. Reliability estimates for each test can be averaged across the ten 'high' and ten 'low' IQ sub-samples, to allow for quick comparison. These figures are shown in table 6.8.

Table 6.8 Average Reliability of Each Test Across Selected Sub-groups of Low and 'High' IQ subjects.

TEST	'LOW'	'HIGH'
Raven's Matrices	.92	.90
Verbal Meaning	.94	.94
Number Facility	.95	.95
Spatial Relations	.81	.80
Figure Grouping	.75	.67
Word Grouping	.79	.70
Perceptual Speed	.95	.95
Word Fluency	.95	.96
Digit Span	.92	.92
Figures of Speech	.78	.81

It can be seen from the above, that for eight of the ten tests in the battery the average reliability estimates obtained for low and high IQ samples are identical or very similar (max difference .03). The most pronounced differences in average reliabilities between 'high' and 'low' IQ samples, are found for the Figure Grouping and Word Grouping tests. In both cases the reliability estimates are lower in the high IQ group. The differences are small, however.

When all of the reliability estimates obtained within the ten 'low' and ten 'high' IQ sub-groups are averaged, the following general picture emerges: 'Low' .88 'High' .86

These figures indicate that there are only negligible general differences in test reliability between range-matched samples above and below mean IQ. This hypothesis was assessed by transforming reliability coefficients within the twenty subsamples to values of Fisher's Z, whereupon corresponding cells in the high and low IQ subgroups were compared using the z test. Values of z pertaining to the differences between the corresponding reliabilities in each pair of 'high' and 'low' IQ samples are shown in table 6.9. For the reasons already stated, the test used to select each pair of samples has not been included in these calculations. Values exceeding plus or minus 1.96 indicate significant differences at the .05 level (two-tailed). (Underlined)

Table 6.9 Differences between reliabilities in 'high' and 'low' ability samples selected by IQ on each of the tests, expressed as values of Fisher's z

(Positive figures indicate that the test's reliability is lower in the 'high' IQ sample).

SAMPLE: PAIR:	RPM LO-HI	VM LO-HI	NF LO-HI	SR LO-HI	FG LO-HI	WG LO-HI	PS LO-HI	WF LO-HI	DS LO-HI	FS LO-HI
TEST										
RPM	*	1.23	1.98	2.79	2.42	2.07	2.59	1.72	.41	.44
VM	-.58	*	1.77	1.13	-1.22	1.12	.59	1.77	-.61	0.00
NF	-.71	2.14	*	.71	-1.5	.71	-.71	2.16	-.76	1.52
SR	-.23	-.95	1.88	*	.23	.47	.96	.49	-.22	2.68
FG	3.43	1.17	3.82	.98	*	.76	2.17	-.99	1.81	2.41
WG	3.06	3.48	3.23	4.01	1.66	*	1.46	3.19	0.00	1.09
PS	-.71	0.00	-.64	0.00	-1.34	0.00	*	-1.36	-1.44	-.71
WF	-.79	-.79	0.00	-.8	-1.71	-.79	-1.52	*	.76	0.00
DS	-.43	.43	.91	.44	.90	-.43	.44	0.00	*	-.44
FS	-1.14	-.91	-.47	-.47	-1.86	-.46	-.94	.23	-.64	*

As can be seen from the above, the degree of difference in reliabilities between high and 'low' IQ samples, varies across sub-groups (selection criteria) and tests. The test showing the largest number of significant differences is *Word Grouping*, for which five out of nine high/low comparisons exceed the critical values. For *Raven's Matrices* and *Figure Grouping*, four of the nine pairs differ significantly. Only two out of five pairs of reliabilities differ for *Number Facility* and only one for *Spatial Relations*. For the remaining four sub-tests, there are no significant differences between reliabilities obtained in the 'high' versus 'low' IQ samples.

The total number of significant differences (17) can be compared to the number expected by chance (4, or 5%) using the chi squared test. This yields an overall value of 37.4, indicating, with >99% confidence (two-tailed), that there are more significant differences in test reliabilities between high and low IQ samples than would be expected by chance.

Although 77 out of a possible 90 differences are <u>not</u> significant, the fact that a greater-than-chance number of reliabilities are lower in the 'high' IQ group, suggests that the correlational data should be adjusted before reaching a final conclusion about g-strength in the two samples.

The original twenty correlation matrices were, therefore, adjusted for test unreliability, using the correction procedures detailed in the last chapter. These adjusted matrices can be found in Appendix 6.5. The amount of common variance in each of the matrices is shown below:

<u>Table 6.10 Strength of g by two estimates in 'high and 'low' ability groups selected by IQ score on each of the ten tests, after correcting for test unreliability within samples</u>

	MEAN CORRELATION		VARIANCE EXPLAINED BY FACTOR 1	
SELECTION TEST	Low IQ	High IQ	Low IQ	High IQ
Raven's Matrices	.25	.25	.35	.35
Verbal Meaning	.29	.28	.38	.37
Number Facility	.25	.25	.34	.35
Spatial Relations	.27	.30	.36	.39
Figure Grouping	.29	.32	.38	.40
Word Grouping	.26	.25	.35	.35
Perceptual Speed	.30	.28	.39	.38
Word Fluency	.30	.34	.40	.43
Digit Span	.43	.39	.50	.46
Figures of Speech	.32	.33	.41	.43

As with the uncorrected data, the strength of Spearman's g is identical or very similar in 'high' and 'low' IQ samples, irrespective of the selection criterion (max. difference: mean r .04, %Var.1st PC .04). Differences in g-strength between samples selected by different tests are, again, closely related to the test's g-loading.

As before, the corresponding correlations in the individual matrices were averaged, to yield overall estimates for 'high' and 'low' IQ samples. These are shown in alternate halves of table 6.11.

Table 6.11

Average of correlations obtained in the 'High' IQ **(top right, bold text)** and 'Low' IQ samples (bottom left, normal text), after correction for test unreliability within sub-samples.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	**.35**	**.44**	**.54**	**.51**	**.47**	**.33**	**.15**	**.18**	**.22**
Verbal Meaning	.45	•	**.30**	**.33**	**.32**	**.45**	**.21**	**.31**	**.19**	**.25**
Number Facility	.44	.36	•	**.33**	**.38**	**.50**	**.48**	**.35**	**.16**	**.34**
Spatial Relations	.47	.42	.37	•	**.51**	**.40**	**.38**	**.20**	**.08**	**.29**
Figure Grouping	.49	.30	.41	.44	•	**.51**	**.25**	**.10**	**.01**	**.21**
Word Grouping	.44	.47	.41	.36	.49	•	**.37**	**.26**	**.15**	**.23**
Perceptual Speech	.30	.25	.51	.37	.29	.41	•	**.26**	**.17**	**.25**
Word Fluency	.22	.32	.36	.21	.14	.29	.34	•	**.07**	**.40**
Digit Span	.23	.13	.19	.09	.07	.12	.17	.09	•	**.20**
Figures of Speech	.18	.19	.23	.26	.14	.23	.25	.34	.16	•

Final estimates of g-strength derived from the above matrices are compared to those found for the uncorrected data in table 6.12.

Table 6.12 Strength of Spearman's g in High versus Low IQ Samples, Averaged Across Selection Criteria, After and Before Correcting for Unreliability.

	AFTER CORRECTION		BEFORE CORRECTION	
	LOW	HIGH	LOW	HIGH
Mean inter-subtest correlation, excluding diagonals.	.30	.30	.25	.26
Percent of total variance explained by the first principal component.	.38	.38	.34	.34

It is clear from these results that, although corrections for test unreliability have the general effect of raising values of r, the strength of the general factor remains equal in the two groups.

In summary, it can be stated that the strength of g is equal in groups selected so as to be in the IQ ranges 65.5-99.0 and 100.5-138.0 on a normally distributed criterion variable, and having means of 88 and 112 relative to a grand sample mean of 100. The differentiation hypothesis is not supported by these findings.

INTERIM DISCUSSION OF RESULTS FOR STUDY 2A.

The finding of equal g-strength in samples above and below mean IQ, apparently conflicts with much of the previous research in this field. The study with which the present analysis can be most easily compared is that of Deary et al. (1996). These investigators also examined samples of below- and above-average IQ, with similar means and standard deviations to those reported above. Using the standardisation data for the Differential Aptitude Test, Deary et al. found a more pervasive g-factor in the lower ability groups, when samples were split by three of the eight scales: Verbal, Numerical, or Spatial Ability. When samples were split by Abstract Reasoning, however, the opposite was found, with a stronger general factor in the higher ability sample. Selection by each of the remaining four tests (Clerical Speed, Mechanical Reasoning, Spelling, Language Usage) did not yield any clear differences in g-strength between 'high' and 'low' IQ groups. Deary et al suggest that their findings offer tentative support for the differentiation hypothesis, but they acknowledge that the results are inconclusive. In view of these mixed findings it would be interesting to re-analyse Deary et al.'s data using matrices of averaged correlations across selection criteria, as in the present study. Unfortunately, the correlation matrices are not supplied in their paper. A close approximation can, however, be derived by averaging the proportionate variance contributions of the first principal component, in samples selected by each of the eight sub-tests. The resultant figures are, for the averaged 'low' IQ sample: 49.8%, and for the averaged 'high' IQ sample: 49.9%. These figures are remarkably similar to one another, bearing-out the results of the current study. The main difference between the two studies, therefore, lies in the diversity of g-strength in samples selected by the various subtests. In this regard it is conceivable that the Differential Aptitude Test, studied by Deary et al., lives up to its name, having a far greater variety of generality or specificity in the sub-tests, than is found in the current battery.

It is less easy to reconcile the current results with the clear pattern of decreasing g-strength with increasing ability level, found by other investigators (e.g. Detterman and Daniel, 1989; Lynn, 1992). One reason for this difference may be that the mean IQs of the samples in the present study, and that of Deary et al., are within the central region of the normal range. As noted by the latter authors, the strongest evidence of ability differentiation tends to have been found when sub-samples at the extremes of the IQ distribution, were included. For example, in one of their two studies, Detterman and Daniel (1989) contrasted mentally retarded young adults with college students. Their second study involved a comparison of five sub-groups from the normative population for the Weschler tests, which incorporates subjects throughout the range of measurable ability. As described in chapter 4, the subject pool for the studies reported in this thesis, while fairly wide in range, has a somewhat higher mean for general intelligence than is the case in the general population. Allowing for temporal increases in intelligence, the average IQ of the whole sample has been estimated to be around 108, relative to the general population. The average IQs of the 'low' and 'high' samples are, therefore, closer to 96 and 120. Although these adjustments do not affect the difference in IQ between samples they bring into question the label 'low IQ' for the below-average sample.

An even more important consideration has to do with the relationship between the direct selector and the incidentally selected variables. As discussed previously, direct selection on one of the variables in a battery does not lead to the same degree of range restriction on the other variables, where they are imperfectly correlated. As already reported, the mean IQs of the 'low' and 'high' ability samples (selected to be within the ranges 62 -> 99.5 and 100.5 -> 138) are 88 and 112, respectively. For the *incidentally selected* variables (i.e. the other nine tests), however, this difference is less pronounced, with the sub-group means tending towards the mean for the entire sample (i.e. 100). This is illustrated below, with selector test means shown in bold.

Table 6.13 (a) Mean IQ on direct (bold) versus incidentally selected variables, where the sample has been selected to be within the IQ range 66 to 99.5 on the former(= 'LOW').

SELECTOR ->	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	**88**	95	94	95	95	95	96	97	97	98
Verbal Meaning	95	**88**	96	95	97	95	97	96	98	98
Number Facility	94	95	**88**	96	96	95	94	95	98	97
Spatial Relations	94	95	95	**88**	95	96	96	97	99	97
Figure Grouping	95	96	96	96	**88**	95	97	98	99	98
Word Grouping	95	95	94	96	95	**88**	96	96	98	97
Perceptual Speed	96	96	93	95	97	95	**88**	95	98	96
Word Fluency	97	97	96	97	99	97	96	**88**	99	96
Digit Span	98	98	98	99	100	99	99	99	**89**	98
Figures of Speech	97	97	97	97	98	97	98	96	98	**88**

Table 6.13 (b) Mean IQ on direct (bold) versus incidentally selected variables, where the sample has been selected to be within the IQ range 100.5 to 138 on the former.(= 'HIGH').

SELECTOR ->	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	**112**	105	106	105	106	106	104	102	103	102
Verbal Meaning	105	**112**	104	104	104	106	103	104	102	102
Number Facility	106	104	**112**	104	104	106	106	104	102	103
Spatial Relations	106	104	104	**112**	105	104	104	103	101	103
Figure Grouping	105	104	104	104	**112**	104	103	101	101	102
Word Grouping	105	105	105	104	105	**112**	104	103	102	103
Perceptual Speed	104	103	107	104	104	105	**112**	104	102	103
Word Fluency	103	104	105	102	101	103	104	**112**	102	104
Digit Span	102	102	102	101	100	102	101	101	**112**	102
Figures of Speech	103	103	103	103	102	102	103	104	102	**112**

In all sub-groups, mean IQs on incidentally selected variables are much closer to the overall sample mean of 100 than is the case for the direct selectors. Across 'low' and 'high' sub-groups, the average IQ levels for the incidentally selected variables are 96 and 103. This difference of 7 points compares to a difference of 24 points for the direct selectors.

This finding highlights the difficulties with this approach to subject selection and raises questions regarding the data reported by other investigators. (With the exception of Deary et al., 1996, all of the studies mentioned report only the sub-group means and ranges on the direct selector variables.) It is particularly important in view of the fact that estimates of g-strength have been wholly, or primarily, derived from intercorrelations between incidentally-selected variables.

Unfortunately, due to the size of the current data pool, it is impossible to create ideal samples which are below or above average in IQ on all ten variables. (Applying the same range restriction to all variables, simultaneously, yields only 14 subjects in the low IQ group and 17 in the high IQ group.) Although imperfect, the method used is, therefore, the most suitable for subject selection in the current data set. Widening the IQ gap between samples will reduce the extent of regression to the mean on the non-selector variables. This approach was taken in the second comparison.

STUDY 2 B: STRENGTH OF G ACROSS THREE RANGES OF ABILITY

AIMS & CONSIDERATIONS

For the reasons discussed in the preceding section, it is important to examine the strength of Spearman's g in subgroups which differ in IQ to the greatest possible extent on incidentally-selected, as well as explicitly-selected variables.

In order to examine the possibility of a curvilinear relationship between IQ and g-strength, it is also necessary to include a sample at the centre of the distribution.

A third consideration, is that there should be enough subjects in each sample for the purposes of estimating test reliabilities and performing principal components analysis. Given the number of variables involved in this study, a sample size of around 100 is desirable (Kline, 1988).

SUB-GROUP SELECTION

The main difficulty in meeting the above criteria involves the number of available subjects. Selecting sub-groups according to a specified IQ range on one test, only restricts the ranges of other test variables to the extent that they are correlated with the selector. It is therefore necessary to restrict the range of the direct selector very severely in order to create groups which differ widely in IQ on the non-selector variables. The more severe this restriction, the fewer subjects will remain in the selected sample. As already noted, sub-samples should ideally contain at least 100 subjects.

Given the size of the current sample (N = 527) it is only possible to select up to three range-matched groups before the size of the lowest and

highest samples drops to an unacceptable level. To create these three groups, the overall sample range of 76 IQ points (62-138) is divided equally into three ranges of 25 points. These ranges are as follows: 'LOW': 62-87, 'MID': 87.5-112.5 and 'HIGH': 113-138. As before, specified groups are selected using each of the IQ variables in turn.
The mean sub-group IQ on these direct selectors, and the corresponding means of the incidentally selected variables, are shown in tables 6.14 a to c. (Figures are rounded to the nearest whole number.)
It can be seen from these tables that, across selection criteria, the mean IQs of the 'LOW', 'MID' and 'HIGH' ability subgroups are 79, 100 and 121 (plus or minus 2 points). With respect to the incidentally selected variables, these figures average 94, 100 and 106, when rounded to the nearest whole number.* The sub-group means produced by the specified range restrictions are fairly consistent across sub-groups. The means of the 'low' and 'high' groups, on selector and non-selector variables, are evenly spaced below and above that of the 'mid' group, for which the mean matches that of the standardisation sample. As with the data for study 2a, a large difference between the mean IQs of the highest and lowest ability groups on the direct selector (42 points) hides a smaller difference in the means of the incidentally selected variables (12 points). Nevertheless, these figures are much more widely spaced than those of the two samples in study 2a, for which the corresponding figures are, for the direct selector: 88 vs 112 (24 points) and, for the incidentally selected variables: 103 vs 96 (7 points).

*It might be thought that these distances are not greatly different to those analysed in study 2a. To increase the distance between the mean IQs of the lowest and highest samples, however, the range of the selector would have to be reduced to a level at which only a very small number of subjects would remain in the samples. For example, range restrictions of 62-82 and 118-138 in the low and high groups would produce means of 75 and 124 on the direct selector but 91 and 108 on the incidentally-selected variables. The size of these samples would, however, be reduced to 58 and 62. (Raven's Matrices used as selector) Such reductions in sample size are not justified by corresponding increases in the gap between the means of the low and high IQ samples.
Histograms provided in appendix 6.6 illustrate the effect of direct selection by Raven's Matrices IQ on the distributions of this variable and the nine incidentally selected variables. It can be seen from these figures that the distributions of the latter remain normal and are therefore suitable for parametric analysis. A similar pattern is evident for samples selected according to the other variables, although figures have not been supplied.

Table 6.14 (a) Mean IQ on direct (bold) versus incidentally selected variables, where the sample has been selected to be within the IQ range 62-87 on the former (='LOW' ABILITY GROUPS)

SELECTOR ->	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
SAMPLE SIZE ->	102	102	97	106	100	103	103	105	100	109
Raven's Matrices	**79**	92	90	92	91	90	92	95	95	97
Verbal Meaning	91	**79**	92	93	94	91	94	94	95	95
Number Facility	91	91	**79**	93	91	92	89	92	95	95
Spatial Relations	90	93	91	**79**	93	93	90	96	97	95
Figure Grouping	91	95	92	92	**79**	92	92	97	97	97
Word Grouping	91	92	91	93	91	**79**	91	94	97	96
Perceptual Speed	92	94	89	91	94	92	**79**	93	96	96
Word Fluency	95	92	92	94	97	93	94	**79**	98	92
Digit Span	95	98	96	99	97	98	96	97	**81**	98
Figures of Speech	96	94	94	94	93	93	95	92	96	**80**

Table 6.14 (b) Mean IQ on direct (bold) versus incidentally selected variables, where the sample has been selected to be within the IQ range 87.5 to 112.5 on the former. (= 'MID' ABILITY GROUPS)

SELECTOR ->	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
SAMPLE SIZE ->	316	322	324	312	336	326	315	318	326	316
Raven's Matrices	**100**	100	100	99	100	100	100	100	100	99
Verbal Meaning	100	**100**	101	100	100	100	100	99	101	100
Number Facility	100	100	**100**	100	101	100	100	100	100	100
Spatial Relations	100	100	101	**100**	100	100	101	100	100	100
Figure Grouping	100	99	100	100	**101**	100	101	100	101	100
Word Grouping	99	99	100	100	100	**101**	100	100	100	99
Perceptual Speed	100	100	100	100	100	100	**100**	101	100	99
Word Fluency	100	100	100	100	100	101	100	**100**	100	100
Digit Span	100	100	100	100	101	100	100	101	**100**	100
Figures of Speech	100	100	100	100	101	101	100	100	100	**100**

Table 6.14 (c). Mean IQ on direct (bold) versus incidentally selected variables, where the sample has been selected to be within the IQ range 113 to 138 on the former.(='HIGH' ABILITY GROUPS).

SELECTOR ->	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
SAMPLE SIZE ->	103	103	98	109	82	98	106	104	101	102
Raven's Matrices	**121**	109	109	111	110	108	108	105	105	106
Verbal Meaning	109	**121**	105	107	107	108	106	108	102	106
Number Facility	110	107	**121**	107	108	109	111	108	104	107
Spatial Relations	109	108	106	**120**	109	108	107	105	102	106
Figure Grouping	109	108	107	107	**122**	108	104	102	101	104
Word Grouping	110	109	108	107	108	**121**	107	106	103	106
Perceptual Speed	108	106	110	107	107	107	**121**	106	104	106
Word Fluency	105	107	107	105	104	104	107	**121**	101	107
Digit Span	105	104	104	103	99	103	104	102	**119**	104
Figures of Speech	105	106	106	106	103	102	106	107	102	**121**

ANALYSIS OF G-STRENGTH ACROSS SAMPLES

As in study 2a, separate matrices of inter-test correlations were calculated for each sub-group, excluding the test which was used as the selection criterion. This produced a total of 30 matrices (10 variables times 3 sub-groups). These are supplied in appendix 6.7.

Estimates of g-strength, in the ten 'low', 'mid'; and 'high' ability sub-samples, are shown in table 6.15.

Table 6.15 Strength of Spearman's g by two estimates, in samples selected according to IQ on each of the ten variables.

	MEAN CORRELATION			VARIANCE EXPLAINED BY FACTOR 1		
SELECTION TEST	LOW	MID	HIGH	LOW	MID	HIGH
Raven's Matrices	.20	.20	.18	.30	.31	.29
Verbal Meaning	.26	.22	.22	.36	.32	.32
Number Facility	.18	.21	.21	.28	.31	.32
Spatial Relations	.22	.24	.23	.31	.33	.33
Figure Grouping	.25	.25	.30	.35	.34	.38
Word Grouping	.24	.23	.20	.34	.33	.30
Perceptual Speed	.24	.23	.19	.33	.33	.30
Word Fluency	.28	.26	.27	.38	.36	.37
Digit Span	.28	.34	.30	.37	.42	.38
Figures of Speech	.24	.28	.32	.35	.37	.41

The above figures reveal no consistent relationship between g- strength and sample IQ. Estimates of g decline with rising IQ when four of the ten tests are used as the selector (RPM, VM, WG, PS). In contrast, g-strength increases with sample IQ when groups are selected by three of the tests (NF, FG, FS). Selection by two of the tests (SR, DS) results in stronger g in the middle group than in the low and high groups, but g is weakest in the middle group when another selection test is used (WF). All of these differences are very small in size, however, with a maximum difference of 0.08 in mean correlation between groups selected by the same variable (FS high vs low).

As in study 2a, the corresponding correlations in the three groups selected by each of the ten variables, were averaged to produce final matrices for 'low', 'mid' and 'high' ability groups. These are shown in tables 6.16 a-c.

TABLE 6.16 (a) MATRIX OF AVERAGED CORRELATIONS ACROSS 'LOW' ABILITY SAMPLES SELECTED ACCORDING TO IQ ON EACH TEST VARIABLE.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.40	.38	.33	.33	.32	.24	.15	.21	.10
Verbal Meaning	.40	•	.32	.32	.21	.37	.21	.31	.13	.13
Number Facility	.38	.32	•	.29	.34	.30	.5	.32	.21	.21
Spatial Relations	.33	.32	.29	•	.32	.26	.3	.17	.07	.17
Figure Grouping	.33	.21	.34	.32	•	.37	.26	.10	.06	.12
Word Grouping	.32	.37	.30	.26	.37	•	.33	.24	.08	.17
Perceptual Speed	.24	.21	.50	.30	.26	.33	•	.31	.16	.16
Word Fluency	.15	.31	.32	.17	.10	.24	.31	•	.09	.25
Digit Span	.21	.13	.21	.07	.06	.08	.16	.09	•	.12
Figures of Speech	.10	.13	.21	.17	.12	.17	.16	.25	.12	•

TABLE 6.16 (b) MATRIX OF AVERAGED CORRELATIONS ACROSS 'MID' ABILITY SAMPLES SELECTED ACCORDING TO IQ ON EACH TEST VARIABLE.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.36	.41	.43	.4	.38	.28	.16	.18	.16
Verbal Meaning	.36	•	.29	.32	.24	.38	.22	.27	.13	.18
Number Facility	.41	.29	•	.30	.31	.38	.47	.34	.14	.22
Spatial Relations	.43	.32	.30	•	.32	.29	.33	.17	.07	.21
Figure Grouping	.40	.24	.31	.32	•	.36	.2	.07	.03	.08
Word Grouping	.38	.38	.38	.29	.36	•	.33	.23	.11	.16
Perceptual Speed	.28	.22	.47	.33	.20	.33	•	.29	.14	.22
Word Fluency	.16	.27	.34	.17	.07	.23	.29	•	.08	.32
Digit Span	.18	.13	.14	.07	.03	.11	.14	.08	•	.13
Figures of Speech	.16	.18	.22	.21	.08	.16	.22	.32	.13	•

TABLE 6.16 (c) MATRIX OF AVERAGED CORRELATIONS ACROSS 'HIGH' ABILITY SAMPLES SELECTED ACCORDING TO IQ ON EACH TEST VARIABLE.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.31	.36	.48	.40	.33	.32	.16	.15	.17
Verbal Meaning	.31	•	.27	.27	.28	.34	.20	.32	.18	.23
Number Facility	.36	.27	•	.28	.28	.39	.42	.31	.13	.29
Spatial Relations	.48	.27	.28	•	.43	.29	.30	.17	.03	.22
Figure Grouping	.40	.28	.28	.43	•	.3	.18	.11	-.04	.21
Word Grouping	.33	.34	.39	.29	.30	•	.28	.18	.09	.14
Perceptual Speed	.32	.20	.42	.30	.18	.28	•	.23	.18	.19
Word Fluency	.16	.32	.31	.17	.11	.18	.23	•	.02	.37
Digit Span	.15	.18	.13	.03	-.04	.09	.18	.02	•	.17
Figures of Speech	.17	.23	.29	.22	.21	.14	.19	.37	.17	•

Estimates of Spearman's g, derived from these averaged matrices are shown in table 6.17.

Table 6.17 Strength of Spearman's g by two estimates, in averaged 'LOW', 'MID' and 'HIGH' IQ samples.

	'LOW'	'MID'	'HIGH'
Mean inter-test correlation, excluding diagonals.	.24	.25	.24
Proportion of variance explained by the first principal component.	.33	.33	.33

These figures are smaller than would have been the case had the ranges of the sub-groups not been restricted. Upwards adjustment is not necessary for the purposes of comparison, however, since the ranges of all three groups have been restricted to an equal degree.

It is clear from table 6.17 that there are no differences in the strength of g between the three groups.

Test reliability in the three samples

Before drawing any firm conclusions regarding g-strength within these three IQ ranges, it is important to to eliminate the potential effects of differential test reliability. For this purpose, test reliabilities were computed within each of the thirty selected sub-groups. Individual reliability estimates are shown in appendix 6.8. Summarised in table 6.18 are the average reliabilities of tests in each of the ten 'low', ten 'mid' and ten 'high' IQ groups.

Table 6.18 Average Test Reliabilities in 'Low', 'Mid' and 'High' IQ Groups Selected by Each Variable.

SUB-GROUP:	LOW	MID	HIGH
SELECTOR VARIABLE			
Raven's Matrices	.85	.87	.84
Verbal Meaning	.88	.86	.84
Number Facility	.88	.85	.82
Spatial Relations	.91	.87	.84
Figure Grouping	.90	.88	.88
Word Grouping	.88	.88	.87
Perceptual Speed	.88	.86	.84
Word Fluency	.89	.86	.85
Digit Span	.87	.88	.86
Figures of Speech	.90	.88	.85

As can be seen from the above, average test reliabilities exceed 0.8 in all sub-groups. For nine of the ten tests, however, there is a small (max .07, SR) but systematic decline in average test reliability with increasing sample IQ.

Next, all thirty correlation matrices were corrected for within-subgroup test unreliability. Individual corrected matrices are given in appendix 6.9. Estimates of g-strength, derived from the corrected matrices, are shown in table 6.19.

Table 6.19 Strength of Spearman's g in samples selected according to IQ on each of the ten variables. Corrected for test unreliability within samples.

	MEAN CORRELATION			VARIANCE EXPLAINED BY FACTOR 1		
SELECTION TEST	LOW	MID	HIGH	LOW	MID	HIGH
Raven's Matrices	.23	.23	.22	.33	.34	.32
Verbal Meaning	.29	.26	.26	.39	.36	.36
Number Facility	.20	.25	.26	.30	.35	.38
Spatial Relations	.25	.27	.28	.34	.36	.38
Figure Grouping	.27	.28	.34	.37	.37	.42
Word Grouping	.28	.26	.23	.37	.35	.33
Perceptual Speed	.27	.27	.23	.36	.37	.34
Word Fluency	.31	.31	.32	.41	.40	.42
Digit Span	.32	.39	.35	.41	.46	.43
Figures of Speech	.27	.31	.39	.38	.41	.48

As with the uncorrected data, the relationship between IQ and g-strength varies with the test used to select the sub-groups. When groups are selected by Raven's Matrices, Verbal Meaning, Word Grouping, or Perceptual Speed, g-strength declines with increasing IQ. The opposite is true when Number Facility, Spatial Relations or Figures of Speech act as selection tests. No clear pattern is evident when the remaining three tests are used to select the groups.

Once again, the corresponding correlations obtained in the ten 'low', ten 'mid' and ten 'high' IQ subgroups were averaged, to produce three final matrices, shown as tables 6.20 a, b. and c.

TABLE 6.20 (a) MATRIX OF AVERAGED CORRELATIONS ACROSS 'LOW' ABILITY SAMPLES AFTER CORRECTING FOR UNRELIABILITY.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.42	.40	.37	.39	.37	.25	.16	.23	.12
Verbal Meaning	.42	•	.34	.36	.24	.42	.22	.32	.14	.15
Number Facility	.40	.34	•	.32	.39	.34	.52	.33	.23	.24
Spatial Relations	.37	.36	.32	•	.40	.31	.33	.20	.08	.21
Figure Grouping	.39	.24	.39	.40	•	.47	.3	.12	.08	.15
Word Grouping	.37	.42	.34	.31	.47	•	.37	.27	.09	.22
Perceptual Speed	.25	.22	.52	.33	.30	.37	•	.33	.18	.19
Word Fluency	.16	.32	.33	.20	.12	.27	.33	•	.10	.28
Digit Span	.23	.14	.23	.08	.08	.09	.18	.10	•	.14
Figures of Speech	.12	.15	.24	.21	.15	.22	.19	.28	.14	•

TABLE 6.20 (b) MATRIX OF AVERAGED CORRELATIONS ACROSS 'MID' ABILITY SAMPLES AFTER CORRECTING FOR UNRELIABILITY

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.39	.44	.50	.49	.45	.30	.17	.20	.18
Verbal Meaning	.39	•	.31	.36	.29	.45	.23	.28	.14	.21
Number Facility	.44	.31	•	.34	.38	.44	.49	.36	.15	.25
Spatial Relations	.50	.36	.34	•	.41	.37	.38	.19	.08	.26
Figure Grouping	.49	.29	.38	.41	•	.48	.24	.08	.03	.11
Word Grouping	.45	.45	.44	.37	.48	•	.39	.27	.13	.21
Perceptual Speed	.30	.23	.49	.38	.24	.39	•	.30	.16	.25
Word Fluency	.17	.28	.36	.19	.08	.27	.30	•	.09	.36
Digit Span	.20	.14	.15	.08	.03	.13	.16	.09	•	.16
Figures of Speech	.18	.21	.25	.26	.11	.21	.25	.36	.16	•

TABLE 6.20 (c) MATRIX OF AVERAGED CORRELATIONS ACROSS 'HIGH' ABILITY SAMPLES AFTER CORRECTING FOR UNRELIABILITY

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.34	.39	.57	.51	.45	.35	.17	.17	.20
Verbal Meaning	.34	•	.28	.31	.35	.43	.21	.33	.20	.26
Number Facility	.39	.28	•	.32	.34	.50	.44	.32	.14	.33
Spatial Relations	.57	.31	.32	•	.58	.40	.34	.19	.03	.27
Figure Grouping	.51	.35	.34	.58	•	.47	.23	.14	-.05	.28
Word Grouping	.45	.43	.50	.40	.47	•	.35	.23	.13	.20
Perceptual Speed	.35	.21	.44	.34	.23	.35	•	.24	.19	.21
Word Fluency	.17	.33	.32	.19	.14	.23	.24	•	.02	.42
Digit Span	.17	.20	.14	.03	-.05	.13	.19	.02	•	.21
Figures of Speech	.20	.26	.33	.27	.28	.20	.21	.42	.21	•

Estimates of g-strength derived from the above matrices are shown below:

Table 6.21 Strength of Spearman's g by two estimates, in averaged 'LOW', 'MID' and 'HIGH' IQ samples, after correcting for test unreliability within subsamples.

	'LOW'	'MID'	'HIGH'
Mean inter-test correlation, excluding diagonals.	.27	.28	.29
Proportion of variance explained by the first principal component.	.35	.37	.38

The above figures show a slight increase in the strength of g with increasing sample IQ, following corrections for unreliability.
The statistical significance of these differences was assessed using Fisher's Z. Comparisons of corresponding correlations obtained in the averaged 'low' versus 'mid' and 'low' versus 'high' IQ groups, revealed no significant differences. When the 'mid' and 'high' groups were compared only one out of the 45* comparisons exceeded the critical value of z (two-tailed), well below the number expected by chance. It must, therefore, be concluded that the strength of Spearman's g is unrelated to ability level, in the samples studied.

*(One half of a 10 by 10 matrix, with the diagonals removed)

SUMMARY AND CONCLUSIONS FOR STUDY 2b.

In summary, the results of study 2b indicate that there is no difference in the strength of Spearman's g between range-matched samples with mean IQs of 79, 100 and 121, even after correcting for differential test reliability. This result is contrary to the hypothesis of ability differentiation with increasing IQ. Neither does it support the hypothesis of a curvilinear relationship between g-strength and ability level.

These findings mirror those obtained in study 2a, where the standardisation sample was split at the mean to form two groups of equal range. In contrast to the latter, the current samples are more widely separated in IQ with respect both to the direct selector and the indidentally selected variables.

In the interim discussion of study 2a it was suggested that the findings might have reflected the fact that the mean IQ of the below-average sample is close to the mean for the general population, when eight points are added to the figures to reflect the current sample's estimated IQ advantage. It was also noted that the correlated variables from which the estimates of g were derived (the incidentally-selected variables), had mean IQ levels just above the population average, when corrected in the same way.

The samples used in study 2b are not subject to these constraints. Even allowing for an 8 point advantage over the general population, the selected range and mean of IQ within the lowest ability sample is well below average. Relative to the population as a whole, the ranges and means of the three samples can be estimated as follows: LOW: 70-95 (mean 87), MID: 95.5-120.5 (mean 108) HIGH:121-146 (mean 129). When the same adjustment is made to the means of the incidentally selected variables, the figures are 100, 108, and 114, respectively. Although the first of these figures cannot be thought of as characterising a 'low' IQ sample, it is essential to point-out that the means and ranges reported by previous investigators* are those of the *direct* selectors.

As discussed previously, it is possible that findings of greater g-strength in low ability samples may be more likely at very low IQ levels, such as those studied by Detterman and Daniel (1989). No subjects who could be classified as mentally retarded attended the school from which the subjects of this study were drawn. Many could, however, be said to fit into the category of slow learners.

With respect to the opposite extreme of the ability range, the means of the 'high' IQ samples of the current study match or exceed those studied by other investigators. The adjusted mean (129) and range (121-146) of the these samples are similar to those of university students. In comparison, Detterman & Daniel's college students had a mean IQ of 115 (WAIS-R), and their 'high-IQ' sample of school children a mean of 122 (WISC-R).
The null findings of the current study cannot, therefore, be attributed merely to a lack of subjects at the extremes of the ability range.

It is the conclusion of this study that there is no evidence to support a hypothesis of ability differentiation with increasing IQ, when adequate precautions against statistical error are made. Since no mentally retarded subjects were included in the sample for this study, the hypothesis that Spearman's g is stronger within the very lowest ability ranges, cannot be refuted.

* With the exception of Deary et al., 1996.

Chapter 7

Studies of the relationship between age, intelligence level, and parameters of information processing speed: Introduction.

Chapters 4-6 document this author's attempts to re-test the two hypotheses of psychometric ability differentiation at higher IQ levels and ages. As discussed in chapter 3, a number of theorists have speculated that psychometric differentiation may be symptomatic of individual and developmental differences in the speed of underlying cognitive processes. In most cases, the efficiency of central processing resources is cited as the major explanatory variable. The studies described in the following chapters will attempt to subject these theories to empirical examination, using measures designed to access parameters of information processing speed.

Problem Definition

Of those models which explain differentiation in terms of cognitive processes, Anderson's theory of Minimal Cognitive Architecture (MCA) is the most elaborate, and has the greatest potential explanatory power. For this reason, most of the studies described in this chapter test hypotheses arising from the MCA model. Anderson's theory, and the associated models of Detterman and Spearman are summarised below:

Anderson (e.g.1992) hypothesises a basic processing mechanism (BPM) which varies between individuals with respect to speed. The BPM determines the efficiency with which either of two specific processors (SP1 and SP2) can function. The two SPs have favoured modes of information-processing: propositional versus analogical, giving rise to a wide range of verbal and visuospatial skills, respectively. The algorithms generated by these processors are executed by the BPM, hence the complexity, capability and variety of specific algorithms is dependent upon the capability (speed) of the BPM. At low levels of BPM efficiency the performance of both SPs is constrained to a similarly low level of functioning, hence Spearman's g is strong. At high levels of BPM efficiency the major constraint on performance shifts to the specific processors, since their full range of algorithms can be utilised. Differences in the relative power of SP1 and SP2 will therefore become more apparent as BPM efficiency increases and it is this which gives rise to the progressive differentiation of measured abilities.

A similar theory has been proposed by Detterman (1987), who speculates that there is a threshold level of central processing speed below which all abilities are constrained to a uniformly low level of functioning. Beyond this threshold, performance on cognitive tasks will better reflect the operation of 'secondary independent processes'.

Both Anderson's and Detterman's models resemble one proposed by Charles Spearman in 1925, using mechanical rather than computational analogies. Spearman likened intelligence to a system containing many 'engines' which are specialised for different tasks but are reliant on an 'energy' source which varies between individuals. At low to moderate levels of energy the main constraint on the performance of the specific engines is the amount of energy which can be directed towards them. At the highest levels of energy, maximum engine capacity has been reached and no additional performance will be gained from further increases (the *law of diminishing returns*). The main determinant of individual differences among high ability people, therefore, is the relative strength or weakness of their specific engines.

Although all of the above models posit specific ability differentiation at higher IQ levels, Anderson and Spearman make different predictions in

regard to age differences. Spearman's view on this matter, is similar to that of Brand (1984), Wilson and Nettelbeck (e.g. Wilson, 1984; Nettelbeck & Wilson, 1985; Wilson & Nettelbeck, 1986) with basic 'energy' or speed hypothesised to increase with ontogenetic development. According to Spearman, the same processes will lead to differentiation with increasing age, as with higher IQ levels.
In contrast, Anderson maintains that fundamental speed of cognitive processing remains unchanged during child development. He argues that reported age differences in cognitive speed mainly reflect increases in 'knowledge'. He maintains that such age differences will be eliminated by better task design, which minimises the potential for strategy use.* In this regard, he hypothesises that age-related performance increments, on measures of cognitive speed, will be greater in the case of *response time* (RT) than *inspection time* (IT), since the former is more prone to extraneous non-speed influences (including 'knowledge') than the latter.

In order to examine Anderson's theory experimentally, it was necessary to identify measures suitable for the assessment of the MCA components. Anderson (1988) hypothesises that the cognitive task most likely to access speed of central processing resources (the BPM) and to be free of strategy use, is the *Inspection Time* test of perceptual intake speed, originally devised by Vickers, Nettlebeck, & Willson (1972). Anderson's (1988) modified IT task was employed in the experiments described in this chapter.

*As discussed in chapter 3, Anderson's model also allows for the existence of BPM-independent modules. These are associated with maturation, in much the same way as Piaget's cognitive stages, and Anderson has hypothesised (Anderson, 1992) that their development may contribute to increases in cognitive speed. However, the type of skills facilitated by these modules (e.g. 'theory of mind') are only of relevance at the very earliest stages of mental development - well before%he ages which will be investigated in this study. For this reason they are not considered relevant to the current investigation.

Since Anderson's verbal-propositional processor (SP1) gives rise to verbal group factor abilities, it can be speculated that its efficiency will be indicated by speed of response on a cognitive task involving verbal stimuli. Posner's letter identification task (Posner & Mitchell, 1967) was selected for this purpose. The efficiency of the spatial-analogical processor (SP2); the hypothesised source of visuospatial group factor abilities; may be similarly indicated by speed of response on a task involving spatial manipulation. Shepard's shape rotation task (Cooper and Shepard,1973) was chosen for this reason. Since the primary unit of measurement in the latter two tasks is response time (RT) it is anticipated that these tasks will be more prone to the effects of learned strategies.

The above tasks are fully described in the relevant method sections. They are introduced here so that hypotheses can be spelled-out clearly.

EXPERIMENTAL HYPOTHESES

Anderson's theory of Minimal Cognitive Architecture gives rise to a number of general and specific hypotheses. Support for some of these has been reported in the literature, whilst for others the evidence is limited. All will be addressed in the current study. They are summarised below as a set of theoretical propositions and predictions.

PROPOSITION:

Individual differences in general intelligence stem from differences in the speed of central information-processing resources. This can be estimated from performance on the inspection-time task and, with lesser confidence, reaction time tasks.[1]

PREDICTIONS:

Individual differences in inspection time (and reaction time) will be negatively correlated with performance on psychometric tests. [HYPOTHESIS 1]

The strength of these correlations will be be associated with the tests' loadings on the general factor. [HYPOTHESIS 2]

1 Performance on RT tasks is more likely to be affected by extraneous factors than performance on IT tasks.

PROPOSITION:

Speed of information-processing does not change during ontogenetic development. Improvements in processing speed with increasing age are a by-product of increases in knowledge, such as a better understanding of task instructions and learned strategies. Performance on reaction time tasks is more likely to be influenced by such factors than is performance on inspection time tasks.

PREDICTIONS:

Age groups having the same mean IQ will not differ significantly with respect to average Inspection Time. [HYPOTHESIS 3]

Average Response Time will differ significantly between IQ-matched age groups, with twelve year olds responding faster than nine year olds. [HYPOTHESIS 4]

The size of correlations between IQ and measures of cognitive processing speed will not differ appreciably between age groups. [HYPOTHESIS 5]

PROPOSITION:

Speed of information-processing is an important determinant of individual (IQ) differences in the strength of Spearman's g, since BPM efficiency underlies the differentiation phenomenon.

PREDICTION:

Measures of cognitive processing speed will correlate more highly with psychometric test variables in samples composed of

a] low versus high IQ subjects [HYPOTHESIS 6]

b] those with slower-than average versus faster than average Inspection Time. [HYPOTHESIS 7]

c] those with slower than-average versus faster than average Response Times [HYPOTHESIS 8]

PROPOSITION:

In Anderson's model the main result of higher BPM speed is to free the potential of the two Specific Processors - SP1 -The verbal-propositional processor and SP2 - the spatial-analogical processor.

PREDICTIONS:

Speed of response to verbal-propositional stimuli will correlate with RT to visuospatial stimuli to a lesser degree in a sample of subjects having faster-than-average Inspection Time (BPM speed) than in a sample with slower-than average Inspection Time. [HYPOTHESIS 9]

Chapter 8:

GENERAL METHODOLOGY FOR THE COGNITIVE STUDIES.

APPARATUS

Computer Hardware

All stimuli were presented on a BBC B computer with a green-on-black raster screen having a refresh rate of 50 Hz (inter-refresh interval = 20 ms). This was connected to a response console (20cm x 10cm) on which were positioned one red button and one yellow button (each 5cm^2). To the centre and front of the console, 2cm equidistant from each response key, was a small grey button (1cm^2) which the subject used to initiate presentation of the Inspection Time stimuli.

Computer Software

1. 'Space Invaders' Inspection Time program, written by Dr. M. Anderson, formerly of Edinburgh University Psychology Department, and used with his permission.

2. Shape Rotation Task
3. Posner letter identification task.

Programs 2. and 3. were down-loaded from the main teaching network of the Psychology Department, with the assistance of Dr. P. Caryl.

DESCRIPTION OF COGNITIVE TASKS AND ADMINISTRATION PROCEDURES

INSPECTION TIME

Inspection time (IT) was tested using a version of the two-lines discrimination task devised by Vickers, Nettlebeck, & Willson (1972). In the classic IT test subjects must make a simple visual discrimination between two lines of markedly different length, joined at the top by a short horizontal bar. Although these lines can be accurately distinguished by almost anyone, the image is only exposed for very brief periods of time. Exposure duration is progressively varied (using either of a number of psychophysical algorithms) to establish the subject's perceptual threshold. At the end of the specified period of exposure a larger 'masking' figure replaces the stimulus figure to delete the image from the iconic store and thereby prevent inspection of the stimulus beyond the chosen period. A subject's 'inspection time' is estimated as the exposure duration at which the discrimination can be made with a specified degree of accuracy (usually between 70% & 97.5%). A fixation cue is presented before the onset of the stimulus, to ensure that the subject's attention is focused to the correct region of the display.

Vickers and later theorists including Nettelbeck (1987) have postulated that Inspection time is a more pure measure of 'mental speed' than reaction time (RT), since the latter contains a response element which is susceptible to extraneous influences. These include individual differences in personality (e.g. tendency to check and re-check before responding), variations in the benefit derived from practice, and the use of idiosyncratic response strategies (e.g. Longstreth, 1984, 1986; Rabbitt, 1985; Detterman, 1987).
Although the classic IT paradigm minimises the number and impact of

such influences, Anderson (1986, 1988, 1992) has argued that comparisons of IT across childhood cohorts may still be confounded by developmental differences in 'knowledge' variables which are independent of processing speed. One such variable is the degree to which subjects understand the response requirements of the task. In the classic IT paradigm the subject must say whether the *left*-or *right*-hand line is the longer of the two. This task may place a greater 'cognitive load' on younger subjects. This potential source of bias can be eliminated by changing the required response to *same* or*different*.
Just as important, argues Anderson, are age differences in concentration and motivation. Using a cue to alert the subject to the forthcoming stimulus, will not necessarily ensure that he or she is paying attention at stimulus onset. This problem can be minimised by allowing subjects to initiate stimulus presentation themselves, by pressing a button. By making the task self-paced, individual differences in the effects of fatigue may also be reduced. Levels of motivation are more likely to be maintained if the task is presented in a more interesting format to that used in the usual IT paradigm, such as a 'computer game'.

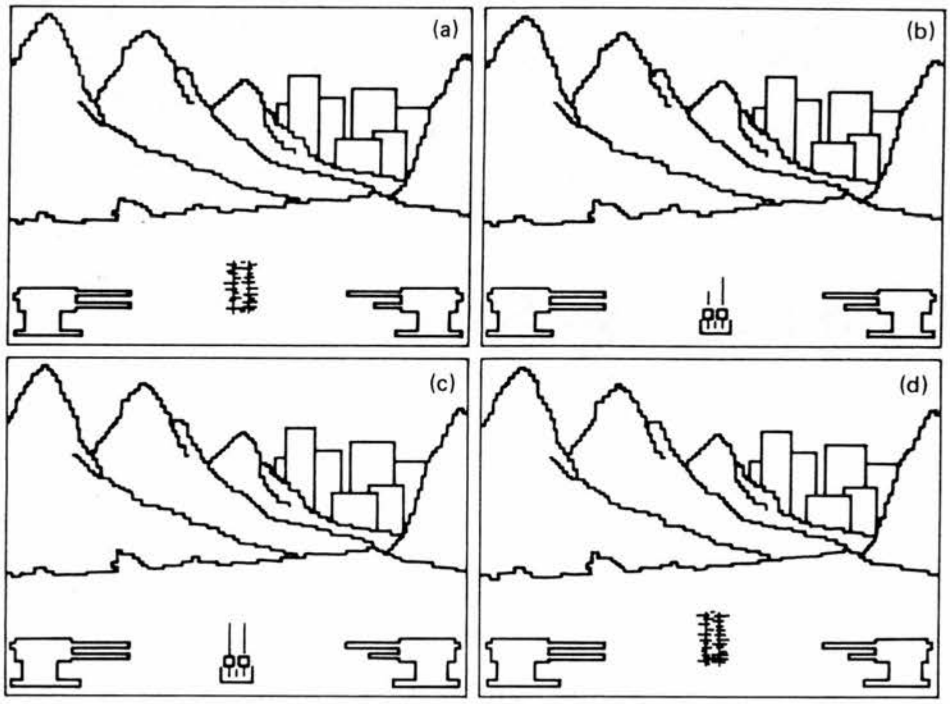
The inspection time paradigm used in the current study was specifically designed to eliminate the effects of extraneous age-related factors. The computer program was devised and written by Dr. Anderson and has been used in several of his published experiments. It is described below:

Anderson's 'Space-Invaders' IT Task.

In Anderson's version of the IT task the basic two-line stimulus is adapted as a "space invader with two antennae". The antennae are either the same length (5mm) or one is shorter (2.5 mm).* * (subtending a visual angle of 0.57 degrees at a viewing distance of 25 cm)
Surrounding the space invader is a monochrome landscape which changes periodically. Each new landscape can be referred to as 'another planet'.

Subjects initiate the presentation of the stimulus figure by pressing a small button at the front of the response console. The task is simply to decide - by pressing one of two coloured response buttons - whether the space invader's antennae are *same* length or *different* lengths. The backward mask consists of a larger block, composed of a hash of short random lines (a "visual patterned mask"). This is referred to as a 'bush' behind which the 'space invader' can 'hide'. To either side of the stimulus are two guns which fire bullets each time a response key is pressed. A correct response or 'hit' is accompanied by the sound of an explosion.

Typical displays are shown below:



Taken from M. Anderson 1990.

IT TASK ADMINISTRATION

Practice Trials

Instructions to subjects were aided by fixed illustrations of 'space invaders' with same-sized versus different-sized 'antennae'. It was explained that the space invader would only be shown for a brief period before 'hiding' and that, sometimes, this period would be so brief that they might not be able to see it at all. It was emphasised that the task was not like an ordinary video game - where the aim is to repond as quickly as possible - and that they could take as much time as they wished to make their decision. (This point was repeated several times during the practice session.) Subjects were then given several practice trials, during which stimuli were exposed for a long enough period for the discrimination to be easily made (between 1.5 s and 400 ms). When subjects had responded accurately to at least 15 stimuli in a row, the practice session was terminated. All subjects had normal or corrected vision.

Psychophysical Procedures Used in Estimating Inspection Time.

After the practice trial, each subject's Inspection Time was estimated using two psychophysical procedures; the Method of Limits (LIMITS) and the Method of Constant Stimuli (MCS). In a slight departure from Anderson's original method, two estimates of MCS IT (rather than only one) were obtained for each subject.

1. METHOD OF LIMITS, with an adaptive staircase

Using this procedure (Anderson's version), the first stimulus is exposed for the relatively long period of 1.5 seconds. Following a correct response, exposure duration is reduced in steps of 100 milliseconds, until the subject begins to make errors. At this point an adaptive staircase algorithm is initiated. Thereafter, correct identifications lead to a further reduction of 20ms, whereas incorrect responses are followed by a 20ms increase. Successive reductions and increases in exposure durations - dependent upon the accuracy of the preceding response - are repeated until the subject's IT threshold is identified. Specifically, the program searches for the exposure duration at which the subject's accuracy rate is approximately 71 per cent (Levitt, 1971). This is interpreted as the subject's inspection time.

[The computer program associated with the LIMITS procedure, only outputs one figure - the estimate of IT at 71% accuracy. This restricts the amount of information available for later analysis]

2. METHOD OF CONSTANT STIMULI (Random order)

The estimate of IT derived from the method of limits is used to set the parameters of the second method, which relies on a larger number of stimulus presentations within a smaller range of exposure durations. For each subject, five exposure durations, 20 ms apart (equivalent to the screen refresh rate), are examined. The central exposure duration (and, consequently the upper and lower levels) is determined by rounding the initial (LIMITS) estimate of IT to the nearest 20 ms. (The computer is programmed to do this automatically.) Subjects are presented with twelve stimuli at each exposure duration, six having same-sized and six different-sized 'antennae'. 'Same' and 'different' stimuli are presented in a random order. Exposure duration is also varied randomly. Where the IT estimate from the LIMITS procedure is less than 20ms, the range of CONSTANT exposures is set to the lowest level (i.e. 20ms -100ms).

From the gradient of exposure duration against average response accuracy, a prediction can be made of the exposure duration needed for 71% accuracy. This figure is taken as the second estimate of IT. (Again, this figure is calculated automatically.)

3. Method of Constant Stimuli: Second Estimate (MCS 2)

The MCS procedure was repeated in the light of concerns, expressed by Anderson (1988), that procedural differences between LIMITS and MCS may lead to error in setting the central MCS exposure.

In contrast to 2, the LIMITS estimate was not used to set the central exposure duration. Instead, one of two estimates was selected, depending on the pattern of response errors across the five exposures of the previous MCS trial. In most cases, accuracy rates varied across the five exposure durations in the expected pattern, with accuracy at the longest duration approaching 100 percent, relatively low accuracy at the shortest duration and around 71 percent accuracy at the central durations. In these cases the first MCS estimate was rounded to the nearest 20 ms and used to set the central exposure for the second MCS estimate. For several subjects, however, the parameters which had been automatically set for the first MCS trials (i.e. following the LIMITS procedure) were too short or too long, resulting in consistently low or high accuracy levels at most or all of the five exposure durations. Where this occurred, the middle exposure for the second MCS procedure was set to a level consistent with the pattern of MCS accuracy rates: In cases where there was a clear change in accuracy rate at some point other than the middle exposure, the exposure producing accuracy rates closest to 71% was nominated as the central level for MCS 2. In cases where all five levels met with consistently high or low accuracy rates, the central exposure for MCS2 was re-set to a level 100ms shorter or longer, respectively, than that set for MCS1. The aim of this was to produce a clear gradient between presentation time and accuracy across the five durations. Where this process did not produce a clear results, the subject was eliminated from further testing. This was necessary in only two cases.

2. RESPONSE TIME TO VISUOSPATIAL STIMULI: SHEPARD'S SHAPE ROTATION TASK.

Speed of mental manipulation was assessed using a version of Shepard's shape rotation task (Cooper and Shepard,1973). The stimuli involved in this task are pairs of two-dimensional shapes, presented on a computer screen. The subject must decide whether the shapes are physically identical, or whether one is a mirror image of the other. The independent variable is the disparity between the spatial orientations of the two shapes, with the orientation of the right-hand figure varied in relation to that of the left-hand figure. It is hypothesised that subjects will use mental imagery to rotate the right hand figure to the upright in order to make the comparison. The dependent variable is the time taken to make the correct response ('same' or 'different') by pressing either one of two buttons. Response time is commonly found to increase as a function of orientation disparity.
Pairs of shapes were presented at four levels of orientation disparity: 0, 60, 120, and 180 degrees. Five individual shapes, each with a mirror image version, were used. At all levels of orientation disparity, both 'same' and 'different' versions of each shape-pair were presented. This resulted in a total of forty trials. New stimuli were presented two seconds after the previous response. To reduce the likelihood of error due to lapses in concentration, each presentation was cued by a high-pitched noise.

Initial test instructions were elaborated with ink outlines and cardboard cutouts, which were used to show subjects how to "twist around" the right-hand figure to see if the pair matched. It was explained that the same rotation process should be conducted "in your mind". It was thought that this demonstration would help to minimise individual differences in strategy use. Next, each subject completed ten computer-presented practice trials, containing equal numbers of 'same' and 'different' shape pairs in a random assortment of orientations. As with the inspection time task, subjects were told to respond 'same' or 'different', by pressing either of two coloured buttons. It was emphasised that they must respond as quickly as possible.

Due to limitations of the test apparatus it was not possible to measure movement time (MT) separately from total reaction time. To miminise the influence of individual differences in MT, subjects were trained to return the index finger of their preferred hand to a specified point on the response console (approximately 3cm from each of the response buttons) immediately after reponding, in anticipation of the next stimulus.

3. REACTION TIME TO VERBAL STIMULI: POSNER'S LETTER DISCRIMINATION TASK.

The letter-matching paradigm developed by Posner (e.g. Posner & Mitchell, 1967) is designed to assess speed of access to long-term lexical memory. Subjects are presented with pairs of letters, where both are of upper or lower case (aa, bb, AA, BB), or where they are mixed (aA, bB, aB, bA). Testing takes place in two phases, in both of which the same stimuli are used. In the first of the two sub-tasks, subjects must decide whether the letters are **physically** identical or physically different. Next, the discrimination task is varied, such that subjects must decide whether the two letters have the same **name** or a different name.
Both tasks require subjects to respond as rapidly as possible, by pressing a 'same' or 'different' response key.
Posner has hypothesised that lexical access speed can be estimated by subtracting the time taken for physical discrimination from the time taken for name identification. For the purposes of the current study, however, mean RT in each condition was considered an appropriate parameter.

Specific administrative procedures: Only the letters A/a and B/b were used in the current experiment. In order to avoid confounding task instructions, one of the Posner subtasks (physical identity or name identity) was administered before the shape rotation task and the other afterwards. The order of presentation was counterbalanced across subjects to eliminate practice effects. In both tests, each of the eight different letter pairs was presented five times - a total of forty stimuli (twenty 'same' and twenty 'different' by each of the two criteria).
Specific test instructions and practice trials were given prior to each sub-test. During practice trials, all combinations of letters were presented twice, in a random order. As with the shape rotation task, subjects were trained to return the finger which they used to respond, to a marked point, equidistant from each response key. In contrast to the shape rotation

task, a new stimulus was presented immediately after each response and without the accompaniment of sound.

ADDITIONAL INFORMATION REGARDING TEST ADMINISTRATION.

All subjects were tested individually. Approximate test administration times are shown below:

Inspection Time:

Instructions and practice: 8 min
LIMITS: 7 min
MCS 1: 3 min
MCS 2: 3 min

Shape Rotation:

Instructions and practice: 5 min
Main task: 5 min

Letter discrimination:

Instructions and practice (total): 6 min
Main tasks (total): 6 min

Total time per subject: 40 minutes (approx.)

SUBJECTS

Sampling Constraints

As already noted, testing took place on an individual basis, with test administration taking around 40 minutes per subject. Subjects were unavailable during breaks, before 9.30 am or after 3.00 pm, and it was not possible to test more than six children per day. For this reason, it was not practical to re-test all of the 527 subjects who had participated in the psychometric studies. Instead, it was decided to draw representative sub-samples from two age groups. For the purposes of the study, these had to be as far apart in age as possible and as wide as possible in IQ range. The school considered the individual testing procedure too arduous for eight year old children. The samples were therefore drawn from the larger cohorts of nine and twelve year olds.

Method of Sub-group Selection

Sub-samples of nine and twelve year-olds were selected on the basis of raw scores on the Raven's Matrices test (RPM). This test was used since it is highly g-loaded and free from ceiling effects and is therefore likely to produce a valid differentiation of subjects in terms of general ability. Efforts were made to ensure that, within each sub-sample, the fullest possible range of RPM raw scores was represented and that these were normally distributed around a mean equal to that of the larger sample. In total, 63 subjects were tested at age nine and 61 at age twelve. Each sample contained roughly equal numbers of boys and girls.

Means, standard deviations and ranges of RPM raw scores at ages nine and twelve are shown below. Also shown are the corresponding figures for the larger samples from which the sub-groups were drawn.

Table 8.1 Mean, Standard Deviation and Range of Raven's Matrices Raw Scores in Sub-samples tested with cognitive tasks, and in the larger samples from which they were drawn.

	AGE NINE				AGE TWELVE			
	Sample size	Mean	S.D.	Range	Sample size	Mean	S.D.	Range
SUBSAMPLE	63	39.1	8.7	39	61	48.2	5.5	26
ORIGINAL SAMPLE	114	39.8	7.4	39	100	48.4	4.7	26

It is clear, from the above, that the selected sub-groups have very similar sample characteristics to those of the larger groups of nine and twelve year olds. Rather than decreasing the standard deviation of RPM scores (as might be expected when a subgroup is selected) s.d.s are slightly increased as a result of the dispersion of a smaller number of cases within the same overall range. As in the larger samples, the range and standard deviation of raw scores is somewhat smaller in the older of the two selected sub-groups. At the time of selecting subjects for individual testing, the main concern was to represent the full range of IQ, so as to maximise the potential range of mental speed. The implications of SD differences were not fully appreciated at that time. Measures to compensate for this difference are described shortly.

Shown in table 8.2 is the mean IQ of each sub-group, and the whole sample, according to published norms for Raven's Matrices (Raven, 1979) and for the Primary Mental Abilities tests (Thurstone, 1963). (No standardisation data is available for the other tests in the psychometric battery, or for the individual Figure Grouping and Word Grouping scales, which are combined as *Reasoning* in the PMA.)

Table 8.2 Mean, Standard Deviation and Range of IQ in Nine and Twelve year old subsamples and in the two groups combined.

	NINE (34 boys, 29 girls)			TWELVE (32 boys, 29 girls)			ALL SUBJECTS (66 boys, 58 girls)		
VARIABLE	MEAN	S.D.	RANGE	MEAN	S.D.	RANGE	MEAN	S.D.	RANGE
RPM IQ (1979)	110.8	18.2	83	115.5	15.2	70	113.1	16.9	87
PMA IQs (1963)									
VERBAL MEANING IQ	118.8	14.2	63	121.1	11.0	54	119.9	12.7	66
NUMBER FACILITY IQ	110.4	8.4	40	103.9	9.2	37	107.2	9.36	40
SPATIAL RELATIONS IQ	108.5	13.5	65	117.7	14.3	60	113.0	14.6	75
REASONING IQ	111.8	13.4	54	117.8	12.4	50	114.7	13.2	54
PERCEPTUAL SPEED IQ	112.1	8.3	35	109.8	8.5	44	111.0	8.43	44
MEAN OF PMA IQS	112.3	11.6	51.4	114.1	10.7	49	113.2	11.7	56

These figures can be compared with those supplied in table chapter 4 for the larger samples tested in the psychometric studies. As with raw RPM scores, average Raven's matrices IQ is approximately the same in subgroups of nine and twelve year olds and the larger samples from which they were drawn.[1] Average Primary Mental Abilities IQ also remains very similar, despite reductions in sample size.[2]
As with the raw scores, the standard deviation of RPM IQ scores is slightly higher in the selected groups than in the larger samples.[3] In contrast, average PMA s.d.[4] is the same in the subsamples and main samples at each age, when figures are rounded to one decimal place.
As with raw scores, the standard deviation of RPM IQ is larger among nine year olds (18.2) than twelve year olds (15.2). The age difference in SD for average PMA IQ is much smaller (11.6 vs 10.7, respectively).

[1] For nine year olds the subgroup mean of 110.8 compares to a mean of 111.8 for the original sample. For twelve year olds the corresponding figures are 115.5 vs 115.2.

[2] For nine year olds the subgroup mean of112.3 is equal to that of the larger group. For twelve year olds average PMA IQ is less than one point higher in the sub-group (114.1 vs 113.4).

[3] (Age Nine: 18.2 vs 15.2. Age Twelve: 15.2 vs 12.6).

[4] (Age Nine: 11.6 vs 12.2. Age Twelve: 10.7 vs 10.7)

Measures to equalise mean IQ in both age groups.

It will have been noted that the two selected age-groups differ with respect to average IQ. Twelve year olds have an advantage of 4.7 points for RPM IQ. Their advantage is less marked for average PMA IQ, at 1.75 points. While these differences are small, it is essential that the groups are matched for IQ, as closely as possible, before drawing conclusions about age differences in parameters of cognitive processing speed. To this end, several cases were eliminated from each of the samples (low scorers aged nine and high scorers aged twelve). In total, six cases were eliminated from the sample aged nine and three from the sample aged 12. Descriptions of the samples remaining are given in the table 8.3.

Table 8.3 Mean, standard deviation and range of IQ in samples of nine and twelve year olds matched, as closely as possible, for mean IQ.
Corresponding figures for the whole sample are also included

	NINE (28 boys, 29 girls)			TWELVE (31 boys, 27 girls)			ALL SUBJECTS (59 boys, 56 girls)		
VARIABLE	MEAN	S.D.	RANGE	MEAN	S.D.	RANGE	MEAN	S.D.	RANGE
* RPM IQ	113.1	16.9	63	114.9	14.6	70	114.0	15.7	70
PMA IQ SCORES									
VERBAL MEANING IQ	120.3	13.7	63	120.9	10.9	51	120.6	12.3	66
NUMBER FACILITY IQ	111.2	7.8	40	103.4	8.8	33	107.3	9.15	38
SPATIAL RELATIONS IQ	109.7	13.5	65	117.9	13.9	60	113.8	14.2	75
REASONING IQ	113.3	12.9	50	117.4	12.3	47	115.4	12.7	50
PERCEPTUAL SPEED IQ	112.6	8.3	35	109.2	8.1	38	110.9	8.3	40
*MEAN OF PMA IQS	113.4	11.2	49.4	113.8	10.8	45.8	113.6	11.3	54

As can be seen from table 8.3, the mean IQs of the nine and twelve year old subgroups are now very close. Although twelve year olds retain a small advantage for mean RPM IQ (1.8 points) the samples are approximately equal for mean PMA IQ. Averaging the two estimates yields an overall mean IQ of 113.3 for nine year olds and 114.4 for twelve year olds. As described in chapter 4, these figures require adjustment to compensate for the rise in IQ since the normative data were collected. Following the same procedures, the mean IQ of the nine and twelve year olds can be estimated to be around 105 and 106, respectively. Such a small difference in mean IQ would not be expected to confound the comparison of cognitive variables in the two age groups.

Significance of remaining age-group differences in IQ variance.

Although the two sub-samples of nine and twelve year olds are now approximately equal in mean IQ, they continue to differ with respect to the standard deviation of Raven's Matrices IQ (16.9 vs 14.6), despite their now equal ranges of raw scores. (After selective removal of cases to produce groups with equal mean IQ the range of raw RPM scores is 26 at both ages and the SD 7.5 and 5.4, respectively.) In contrast, average s.d. across the five PMA IQs is very similar in both age groups (11.2 vs 10.8). Differences in s.d. of this magnitude have not been considered problematic in previous studies comparing age groups with respect to IT-IQ correlations (e.g. Anderson, 1988; Wilson *et al* 1992). Nevertheless, it is important to establish whether the differences are statistically significant. Effective equality of variances was evaluated using the F-test. For IQ on Raven's Matrices and four of the five PMA tests, the age difference was not significant at the 10 percent level, using a two-tailed test. For the Verbal Meaning test, IQ variances were significantly different at the 10% level but not at the 2% level (two-tailed). These results indicate that the size of correlations can be meaningfully compared across age groups.

Chapter 9

Results for Study 3:
Testing Anderson's Model Using Inspection Time as an Indicator of the Efficiency of the Basic Processing Mechanism.

Inspection Time Results for the Combined Sample (N=115)

Descriptive statistics:

As already described, three estimates of Inspection Time (IT) were obtained for each subject and these were averaged to give a composite score. The mean, standard deviation and range of raw IT, in the combined sample, are shown in table 9.1. Coefficients of skewness and kurtosis are also supplied.

Table 9.1 Mean, Standard Deviation, Range and Normality of Raw IT in the Combined Sample of Nine and Twelve Year Olds [N=115]

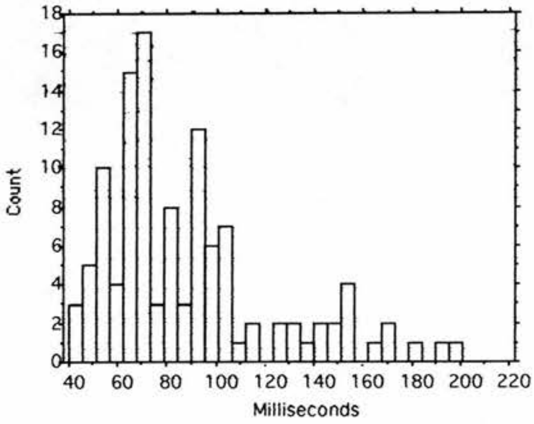
	MEAN	S.D	RANGE	Skewness	Kurtosis
LIMITS	88.6	34.7	156.7 (40.0 - 196.7)	1.21 ($p<.10$)	0.89 (n.s.)
MCS 1	64.6	21.4	151.3 (6.0 - 157.3)	0.95 (n.s.)	3.96 ($p<.01$)
MCS2	58.5	16.9	96.9 (0.8 - 97.7)	-0.64 ($p<.10$)	1.09 ($p<.05$)
Mean IT	70.5	17.4	90.0 (32.1 - 123.2)	0.55 (n.s.)	0.45 (n.s.)

The distributions of all variables meet good criteria for normality, with none having skewness coefficients which are significant at the five percent level. The distribution of LIMITS is moderately skewed ($p<.10$) and MCS1 shows the greatest degree of kurtosis ($p<.01$). The mean, standard deviation and range of IT decline with each successive testing. This may result from practice or it may be the consequence of different psychophysical methods in the case of LIMITS and MCS.
Frequency histograms, illustrating the distributions of raw IT scores are supplied overleaf.

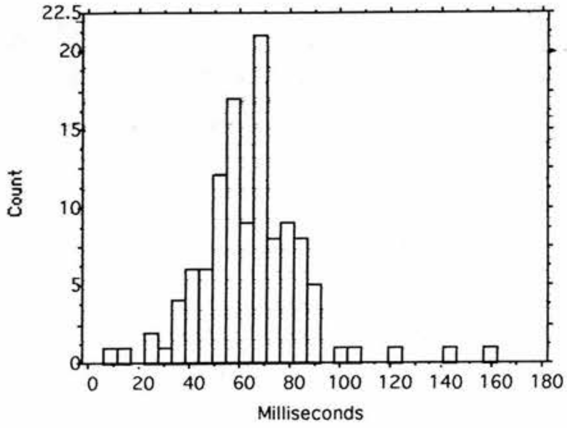
FIGURES 9.1 a.-d.

DISTRIBUTION OF INSPECTION TIME IN THE COMBINED SAMPLE [N=115], FOR EACH OF THE THREE ESTIMATES AND FOR THE MEAN OF THE THREE.

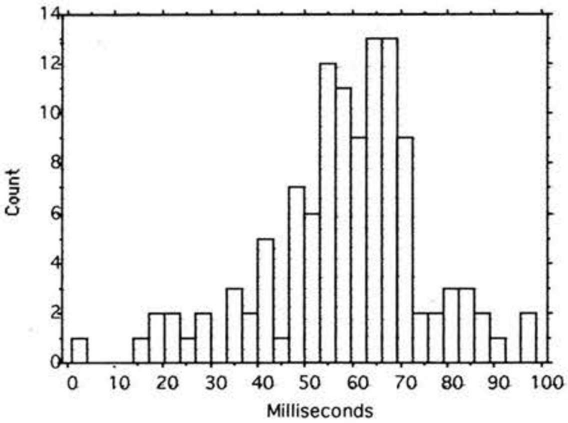
9.1 (a) IT LIMITS



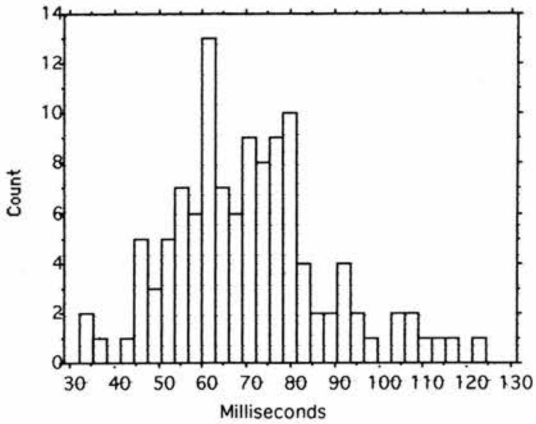
9.1 (b) IT MCS1



9.1 (c) IT MCS 2



9.1 (d) MEAN OF 3 IT ESTIMATES



Correlations between Inspection Time and IQ in the Combined Sample

Shown in table 9.2 are the correlations between the three Inspection Time measures and each of the ten tests in the pencil-and-paper battery. To eliminate the influence of age variance, all variables were standardised within age groups. This was done via normalisation, which enhances the comparability of correlations across variables by removing differences in their distributions. The distributions of all normalised variables have a mean of 0, a standard deviation of 1 and a range of -2.3 to +2.3 Z-scores.[1] For ease of visual comparison, the non-verbal and verbal tests have been grouped separately in table 9.2, as have the two numerical tests. Separate correlations have been calculated for a composite IT variable, derived from the average of the three raw ITs. As with the individual estimates, average IT was standardised within age groups to the same range, i.e -2.3 to +2.3 Z scores (= "Comp SIT").

Table 9.2 Correlation between age-standardised Inspection Time estimates and IQ scores, where all variables have the same range in both age groups and in the sample as a whole.

	IT LIMITS	IT MCS 1	IT MCS2	Comp SIT
Raven's Matrices	-.20	-.14	-.38	-.31
Spatial Relations	-.27	-.22	-.38	-.38
Figure Grouping	-.07	-.02	-.28	-.13
Perceptual Speed	-.24	-.12	-.12	-.24
Verbal Meaning	-.09	-.19	-.07	-.12
Word Grouping	.04	.03	-.01	.05
Word Fluency	-.01	-.13	-.07	-.07
Figures of Speech	-.16	-.14	-.12	-.18
Number Facility	-.12	-.02	-.11	-.10
Digit Span	-.10	-.09	-.09	-.12
Average	-.12	-.10	-.16	-.16

[Critical values of r for a one-tailed test = .16 (p<.05), .19 (p<.025), .25 (p<.005)]

[1] Normalisation is unlikely to distort the true scores. Correlations between raw and standardised scores, calculated by this method, are high: LIMITS r=.92, MCS1 r=.93, MCS2 r=.97, Mean/Composite IT r= .96.

Although the correlations between IT and IQ are generally low, they are almost all in the predicted (negative) direction, offering support for hypothesis 1.

Across IT estimates, correlations with IQ are highest for non-verbal tests of intelligence, particularly Raven's Matrices and Spatial Relations. Correlations of IT with verbal and numerical test scores are much lower, in line with other findings reported in the literature (e.g. Nettelbeck, 1987; Kranzler & Jensen, 1989).

The single IT estimate which, on average, correlates to the greatest degree with IQ is MCS2, the third in the sequence of measurements. It might be expected that the composite IT variable would show higher correlations with psychometric abilities than either of the three individual estimates. As can be seen from table 9.2, however, the composite IT variable correlates more highly with only two of the pencil-and-paper tests (Figures of Speech and Digit Span). The variable which has the highest correlation with the estimate of general intelligence (Ravens Matrices) is MCS 2. This brings into question the advantage of using the composite measure in preference to the individual ones (notably MCS2), when investigating the association between IT and IQ.

Reliability of Inspection Time in the Combined Sample.

An indication of the reliability of each IT estimate is given by the inter-correlations between the three measures, shown below:

Table 9.3 Correlations between standardised IT estimates in the combined sample [N=115]

	LIMITS	MCS1	MCS2
LIMITS	1.00	.23	.29
MCS1	.23	1.00	.36
MCS2	.29	.36	1.00

The correlations between IT estimates, shown in table 9.3, are rather lower than those typically reported in studies of inspection time (e.g. Nettelbeck, 1987). This may explain the failure of the composite IT measure to have uniformly higher correlations with IQ than either of the three individual estimates. Unfortunately no equivalent figures are provided by Anderson (1988) for the purpose of comparison.

An estimate of the reliability of each measure can be obtained from the average of its correlations with the other two measures.[1] Thus estimated, the reliabilities of the three IT tests are as follows:

LIMITS = .26 MCS1 = .29 MCS2 = .33

These figures suggest that IT test reliability is highest where IT is measured by the second application of the method of constant stimuli (MCS2), an inference which would be consistent with the finding that, of the three IT estimates, this is this one which correlates to the greatest extent with Ravens Matrices. The figures also suggest that the reliability of the measures increases over the testing sequence, although this conclusion cannot be drawn from the available evidence.

[1] An estimate of internal consistency reliability would have been preferred. Unfortunately, adequate data were not available for this purpose.

INSPECTION TIME AT AGES NINE AND TWELVE

Descriptive statistics:

The means, standard deviations and ranges of IT in the IQ-matched groups of nine and twelve year olds, are shown in table 9.4. Frequency histograms are supplied overleaf (figs 9.2 a-f).

Table 9.4 Inspection Time (ms) in nine and twelve year old samples matched for mean IQ

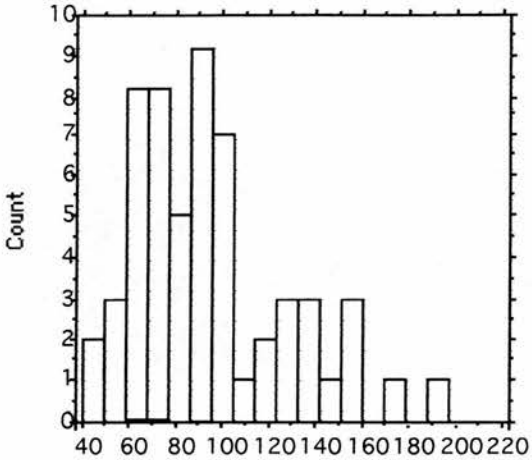
Estimate	NINE YEAR OLDS (N=57)			TWELVE YEAR OLDS (N=58)		
	MEAN	S.D.	RANGE	MEAN	S.D.	RANGE
1. LIMITS	94.5	33.5	156.7 (40.0-196.7)	82.8	35.1	146.7 (43.3-190.0)
2. MCS 1	68.5	24.2	151.3 (6.0-157.3)	60.3	17.4	92.6 (13.8 - 106.4)
3. MCS 2	60.9	15.7	78.8 (18.0-96.8)	56.2	17.9	96.9 (0.8 - 97.7)
COMP IT	**74.6**	16.8	80.0 (43.11-123.15)	**66.4**	17.3	80.91 (32.1-113.0)

As table 9.4 shows, mean IT is longer in the sample of nine year olds, using the composite measure and each of the three individual estimates. The degree to which age groups differ declines steadily across the successive testing procedures, however. The statistical significance of the age differences was evaluated using t-tests (one tailed). For the composite measure the difference is significant at the .005 level ($t=2.59$). For IT estimated with the method of limits the age difference is significant at the .05 level ($t=1.82$). For the first estimate by the method of constant stimuli, twelve year olds are, again, significantly faster ($t=2.09$, $p<.025$). The difference between the means of nine and twelve year olds fails to reach significance for the second MCS estimate, however ($t=1.51$ $p> .05$). These mixed results cannot be interpreted as offering clear support for Anderson's hypothesis that inspection time does not change with age (hypothesis 3).

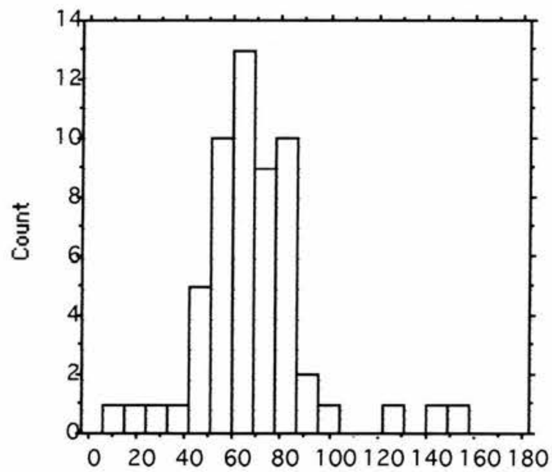
Figures 9.2 a - f. Distribution of Inspection Time (in milliseconds) using three estimates

NINE YEAR OLDS (N=57)

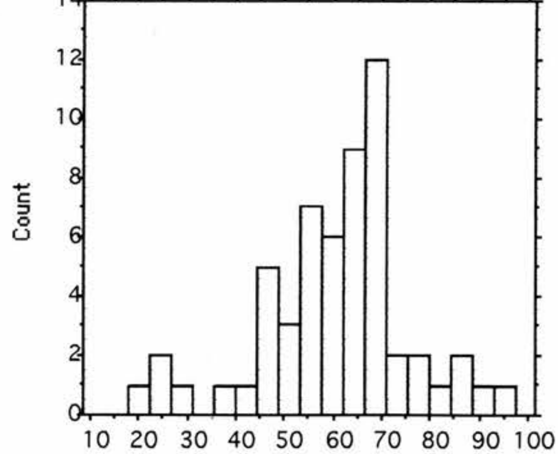
(9.2 a) LIMITS



(9.2 b) MCS 1

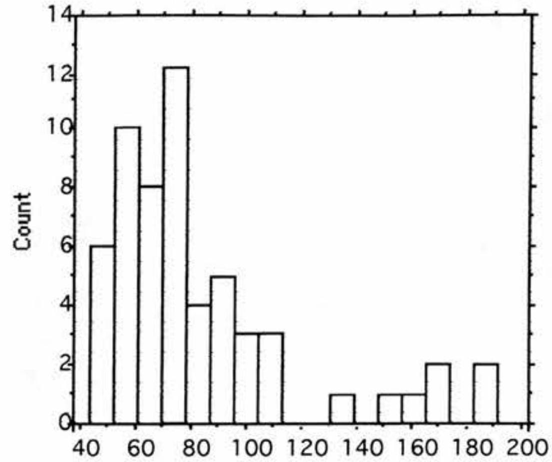


(9.2 c) MCS 2

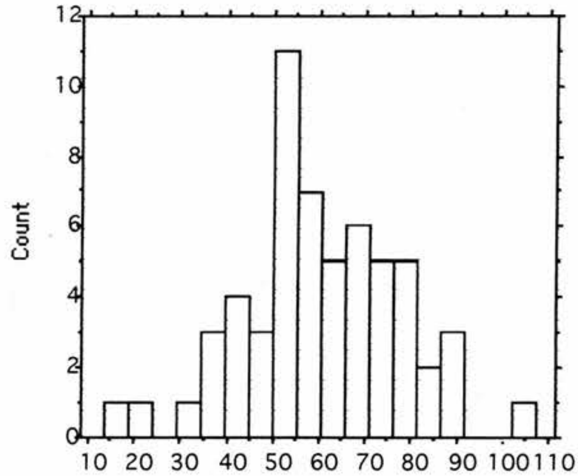


TWELVE YEAR OLDS (N=58)

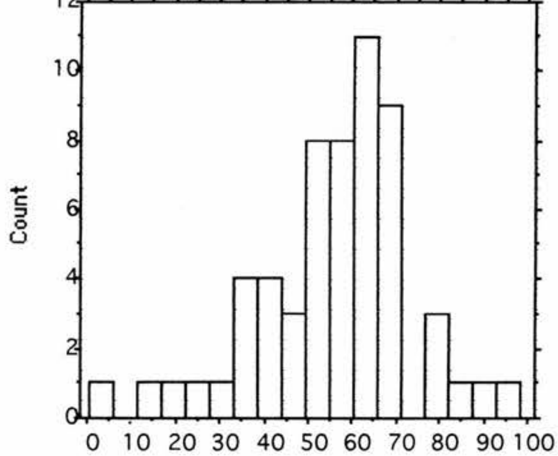
(9.2 d) LIMITS



(9.2 e) MCS 1



(9.2 f) MCS 2



In both age groups, as in the full sample, mean IT declines with each successive testing. The difference is most pronounced between LIMITS and MCS1, suggesting that it may be due to variations in psychophysical procedure. The smaller decline in IT between MCS 1 and MCS 2 may indicate improvement as a result of practice.

The standard deviation of IT differs between age groups, although not always in the same direction. In contrast to IQ scores, twelve year olds have a slightly higher s.d. than nine year olds for LIMITS and MCS2. The age-group difference in variance is only significant for MCS1, with twelve year olds having a narrower distribution ($F=1.94$, $p<.01$, two-tailed).

Reliability of Inspection Time Measures in the Nine and Twelve Year Old Samples

Table 9.5 shows intercorrelations between the three IT estimates in each age group and in the combined sample. The comparability of the correlations is improved by standardising the IT variables within age groups via the process of normalisation. This eliminates differences in the range of IT between measures and between age groups.[1] For comparison, the corresponding correlations between raw IT variables are shown in brackets.

Table 9.5 Correlations between standardised [and raw] inspection time variables in nine and twelve year-olds and in the two samples combined.

	AGE 9 (N=57)		AGE 12 (N=58)		ALL SUBJECTS (N=115)	
	LIMITS	MCS1	LIMITS	MCS1	LIMITS	MCS1
MCS1	.10 [.01]	****	.36 [.24]	****	.23 [.14]	****
MCS2	.30 [.27]	.32 [.35]	.28 [.22]	.39 [.40]	.29 [.26]	.36 [.38]

[Critical values of r for a one-tailed test = .21(p<.05), .25 (p<.025), .32 (p<.005)]

[1] The distributions of all normalised variables have a mean of 0, a standard deviation of 1 and a range of -2.3 to +2.3 Z-scores.

The two IT estimates which correlate with each other to the greatest degree, are MCS 1 and MCS 2, with the correlation slightly higher in the older sample. Interestingly, raw LIMITS IT does not correlate significantly with MCS1 in nine year olds, although it does so at age twelve. Raw LIMITS and MCS2 are significantly related in both nine and twelve year old samples. Table 9.6 shows the average of the correlations between each IT estimate and the other two.

Table 9.6 Estimated reliability of each IT measure in the two age groups and in the sample as a whole, calculated using standardised (and raw) IT variables.

	AGE 9	AGE 12	ALL SUBJECTS
IT LIMITS	.20 [.14]	.32 [.23]	.26 [.20]
IT MCS1	.21 [.18]	.38 [.31]	.29 [.26]
IT MCS2	.34 [.31]	.34 [.31]	.33 [.32]

As in the combined sample, there is an increase in the reliability of IT with successive testing procedures in the nine year old sample. A similar pattern is evident amongst the twelve year olds, although in this group the reliability of (raw) IT is the same for the second (MCS1) and third (MCS2) measures.

Reliability estimates in the two age groups were compared using Fisher's z test.[1] In no case is the difference statistically significant, indicating that differences in measurement reliability are unlikely to be a serious source of bias when comparing the correlations of IT with other variables, across age groups. In fact, the reliability of the final IT (MCS2) estimate is equal in the samples of nine and twelve year olds. These figures suggest that MCS2 is the most reliable measure of the three, indicating that MCS2 is the most suitable for comparing IT-IQ correlations across age groups. The composite IT variable will also capitalise on the strengths of each individual estimate.

[1] [Values of z indicating the difference between age groups in the estimated reliability of each IT measure are as follows: LIMITS z = .68 n.s., MCS1 z = .98 n.s., MCS2 z = 0]

The pattern of IT results obtained using the three estimation methods warrants further discussion, before drawing conclusions about the age difference in inspection time. Of particular note is the apparent decline in mean IT with each successive testing. As already noted, the difference in IT between estimates, is greatest for LIMITS compared to MCS1, suggesting that it is due to differences in psychophysical procedure. Nonetheless, the faster IT for MCS2 than MCS1 also suggests that practice may be a contributory factor. Practice may reduce the influence of extraneous factors adding to the variability of scores (i.e. error variance) resulting in a more 'pure' estimate of cognitive speed. The reduction, with successive testings, of the standard deviation and standard error of IT, coupled with increasing test reliability, is consistent with this explanation. In this regard, it is noteworthy that the age difference in IT reduces with each new testing procedure, declining to non-significance by the third testing (MCS2). This finding would support the hypothesis that age groups initially differ with respect to their susceptibility to extraneous influences but that practice reduces these effects, with younger subjects benefitting more than older ones, but over a longer period of time. Nettelbeck and Vita (1994) have found results concordant with this explanation and Anderson (1988) has observed that "the ability to benefit from practice is related to MA and not to IQ". Also consistent with this suggestion is the current finding of similar standard deviations for MCS 1 and MCS 2 in twelve year olds, but an apparently continuing drop in s.d. for the nine year olds.

It should be noted that Nettelbeck & Vita (1994) hypothesise that practice may improve performance as a result of strategy aquisition. In other words, they suggest that practice may produce a <u>less</u> pure measure of IT. This explanation does not accord with the current finding of reduced IT standard deviations at later testing sessions, however. Nevertheless, only three estimates were taken in this experiment, compared to over twenty by Nettelbeck & Vita. It may be the case that if practice is continued beyond the level of peak performance, strategy use may begin to play a more important role in accounting for individual differences in performance. Nettelbeck and Vita's finding, that IT-IQ correlations decline to non-significance after extensive practice, is consistent with this explanation.

Differences in mean IT across the three testings may also be due, in part, to unintentional procedural factors. In both age groups the largest difference is between IT estimated from the method of limits and the first MCS procedure, despite the fact that the LIMITS estimate is used to set the parameters for MCS1. A similar finding was obtained by Anderson (1988), using the same computer programs. Although he attributed the difference partly to practice effects, Anderson acknowledged the possibility that "procedural differences between the method of limits and the method of constant stimuli [may] produce overestimates of IT using the former". He also conceded that "the assumptions underlying the calculation of expected response accuracy in the constant procedure [may be] erroneous". Since there were only five alternative exposure durations for each MCS session, 20ms apart, an underestimate or overestimate of LIMITS IT of only 40ms would compromise the capacity of the first MCS program to produce a clear slope of accuracy on speed. It was for this reason that a second MCS testing session was used, whereby the central SOA was set to the MCS1 exposure which had produced accuracy rates closest to 71%. In view of this problem it is likely that MCS2 is a more accurate estimate of IT.

In summary, Anderson's hypothesis that cognitive speed "does not develop" is only supported in the case of MCS2, for which the age difference in mean IT is not statistically significant. There are significant age differences in mean IT by the first two estimates, however. It has been argued that MCS 2 is likely to be the most valid and reliable estimate of the three, nonetheless, hypothesis 3 cannot be accepted conclusively.

CORRELATIONS BETWEEN IT AND IQ IN NINE VERSUS TWELVE YEAR OLDS.

As already noted, the standard deviation of most IQ and IT variables is not significantly different in the samples of nine and twelve year olds. Nevertheless, it is important to eliminate the effects of these differences as far as possible, when comparing the two age groups with respect to the size of their IT-IQ correlations. It is also useful to bring all tests - including those for which no published IQ conversion tables exist - to a common distribution. To achieve these goals, raw scores on the ten psychometric tests and the three IT estimates were standardised within each age-group using the normalisation procedure described in chapter 5 In all cases the resultant z-scores are distributed around a mean of 0, with a standard deviation of 1.0 and a range of -2.3 to +2.3.* In both samples, IQ scores derived from published conversion tables and the normalised scores for the same tests (SIQ) correlate above 0.93, although the distributions are wider and more normal in the latter. A composite IT score was derived by averaging the three separate IT estimates and standardising the product to the same range.

Correlations between age standardised IQ variables and the four estimates of IT (composite IT included) are shown, for each age group, in table 9.7. For convenience, the four non-verbal tests involving an element of visuo-spatial comparison have been grouped separately from the four verbal tests. Correlations of IT with Digit Span and Number Facility are given after these.

*As described in chapter 5, restrictions in sample size and raw score range can contrain the potential distribution of normalised Z-scores. Where the MINITAB statistics package failed to extend the normalised range to the specified levels, this was done manually. Only minor manual adjustments were required, however, since most of the normalised variables extended between -2.3 and +2.3. In no case did such an adjustment change the correlation between normalised IQ and IT by more than 0.02.

Table 9.7 Correlations between IQ (SIQ) and Inspection Time (SIT) in samples aged nine and twelve years.

	AGE NINE [N=57]				AGE TWELVE [N=58]			
	LIM	MCS1	MCS 2	Comp IT	LIM	MCS1	MCS 2	Comp IT
Raven's Matrices	-.19	-.12	-.35	-.29	-.21	-.15	-.41	-.33
Spatial Relations	-.28	-.22	-.41	-.41	-.27	-.22	-.35	-.34
Figure Grouping	-.19	-.04	-.32	-.24	.04	.01	-.24	-.03
Perceptual Speed	-.23	-.05	-.07	-.22	-.25	-.20	-.18	-.26
Mean Non-Verbal	-.22	-.11	-.29	-.29	-.17	-.14	-.30	-.24
Verbal Meaning	-.16	-.13	-.14	-.17	-.02	-.24	-.01	-.08
Word Grouping	-.08	.02	-.09	-.06	.14	.03	.08	.15
Word Fluency	-.04	-.23	-.02	-.07	-.05	-.03	-.12	-.08
Figs of Speech	-.07	-.22	-.09	-.17	-.25	-.07	-.16	-.20
Mean Verbal	-.09	-.14	-.09	-.12	-.05	-.08	-.05	-.05
Number Facility	-.14	.08	-.11	-.06	-.09	-.12	-.12	-.14
Digit Span	-.11	-.21	-.14	-.16	-.09	.01	-.04	-.08
Mean (All tests)	-.15	-.11	-.14	-.18	-.11	-.10	-.16	-.14

[Critical values of r for a one-tailed test = .21(p<.05); .25 (p<.025); .32 (p<.0005)]

In both age groups, as in the combined sample, the vast majority of correlations between IT and IQ are in the expected (negative) direction, supporting hypothesis 1. The highest correlations are found for MCS 2 with Raven's Matrices and Spatial Relations (RPM & MCS2 at age nine r= -.35, at age twelve r= -.41. SR at age nine: r= -.41, age twelve r= -.35). In many cases, the correlation of IQ with MCS2 is higher than or equal to that with Composite IT.

In both age groups, IT correlates to a greater degree with scores on non-verbal tests than with scores on verbal tests. Illustrative of this difference is the higher correlation of IT with Figure Grouping (FG) than with Word Grouping (WG), despite the fact that these two tests were designed to be highly correlated with each other. In fact, IT and WG correlate close to zero in both samples, compared to correlations of-.32 and -.24 for FG with MCS2 at ages nine and twelve. For the non-verbal tests, nineteen of the thirty-two possible IT-IQ correlations (both samples) are statistically significant, compared to only four out of thirty-two correlations between verbal IQ and IT.

TESTS' G-LOADINGS AND THEIR CORRELATIONS WITH IT.

The hypothesis that the strength of the IT-IQ association is related to the ability test's loading on general intelligence, was examined next. Subjecting the ten psychometric variables to Principal Components Analysis yields the following loadings on factor one:

Table 9.8 Individual Tests' Loadings on the First Unrotated Principal Component.

	AGE NINE	AGE TWELVE
Raven's Matrices	.76	.73
Spatial Relations	.58	.58
Figure Grouping	.70	.60
Perceptual Speed	.57	.70
Verbal Meaning	.60	.61
Word Grouping	.71	.59
Word Fluency	.37	.65
Figures of Speech	.54	.58
Number Facility	.78	.59
Digit Span	.15	.38

The association between the g-loading of each test and and the size of its (negative) correlations with IT, was examined for each IT variable, in each sample, using Spearman Correlations. Results are shown in table 9.9.

Table 9.9 Association between tests' g-loadings and the size of their correlations with IT (Spearman's Rho).

	Age Nine	Age Twelve	Full Sample
LIMITS	-.31	.13	-.16
MCS1	.77	- .30	.47
* MCS2	- .34	-.28	-.43
Comp IT	.01	-.12	-.16

In the case of the most reliable IT estimate - MCS2 - there is a clear association between the g-loading of psychometric variables and the size of their (negative) correlations with IT. For the two less reliable measures, the association is less clear, differing in direction in the two age groups. The association is lowest for the composite IT variable, undoubtedly because the negative and positive associations for the individual estimates cancel each other out.

These results only offer partial support for hypothesis 2.

It should be noted that the results also indicate that test characteristics other than g-loading are important in predicting the size of the IT-IQ correlation. For example, as can be seen in table 9.7, in the combined sample Figure Grouping and Word Grouping have approximately the same loadings on the first unrotated principal component (.65 & .64 respectively) yet their correlations with IT (MCS2) differ markedly (FG & IT -.25, WG & IT 0.00). Likewise, although Raven's Matrices is more highly g-loaded than Spatial Relations (.74 vs .57) the tests' correlations with IT (MCS2) are approximately equal (-.37, -.36).

Relative magnitude of IT-IQ correlations in the two age groups

The next hypothesis to be examined, predicts that the size of the correlation between IT and IQ will be the same in samples aged nine and twelve. (i.e. Anderson's theory that BPM speed does not change with age.)

To check whether any of the forty IT-IQ correlations (10 tests X 4 estimates of IT) are significantly different in the two age groups, all values of r were transformed to Fisher's Z, then compared across age groups using Fisher's z-test. Values of z are shown below:

Table 9.10 Values of z Indicating Differences between Nine and Twelve Year Olds in the Size of IT-IQ Correlations.

	LIMITS	MCS 1	MCS 2	Comp IT
Raven's Matrices	.11	.16	.37	.23
Spatial Relations	-.06	0.00	-.37	-.43
Figure Grouping	-1.21	-.26	-.45	-1.12
Perceptual Speed	.11	.80	.58	.22
Verbal Meaning	-.74	.60	-.68	-.48
Word Grouping	-1.15	-.05	-.89	-1.10
Word Fluency	.47	-1.07	.52	.05
Figures of Speech	.97	-.80	.37	.16
Number Facility	-.26	1.05	.05	.42
Digit Span	-.11	-1.16	-.53	-.42

[Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

None of the forty correlations between IT and IQ are significantly different in the two age groups. This result is in the direction predicted by hypothesis 5.

N.B. It should be noted that the age difference in the size of the IT-IQ correlation is smallest in the case of MCS2 which, it has been argued, is the most reliable and valid of the three estimates.

CORRELATIONS OF IT WITH IQ VERSUS MENTAL AGE.

It has already been reported that the two age groups do not differ significantly in inspection time, when IT is estimated according to the most reliable method (MCS2). This estimate of IT does correlate negatively with most IQ variables however, particularly those with high g-loadings and a strong visuo-spatial component. (This is also true for LIMITS and MCS1.) It can be concluded from this that IQ is a better predictor of IT than age is. In fact, raw IT (MCS2) correlates with CA at -.14, compared to a correlation of -.38 with Raven's Matrices IQ and -.24 with PMA IQ.[1] Nonetheless, other investigators (e.g. Anderson, 1988; Wilson et al, 1992) have made much of the difference between the correlation of IT with *MA* (as an indicator of age-related differences) as compared to IQ, which behoves this researcher to consider this comparison. When contrasting IT-IQ and IT-MA correlations, however, previous researchers have failed to take note of the fact that the absolute values of the correlation coefficients are not directly comparable, since the range of raw scores which exists across age groups (as in the determination of MA) is wider than that which exists within age groups (used to determine IQ). The greater 'true score' variance within MA will tend to artificially enhance the size of correlations between MA and IT, even though the ability-IT relationship may be equally strong, in 'true' terms, within age groups. This problem can be eliminated by normalising raw test scores in the whole sample (MA) and raw scores within age groups (IQ) to the same distribution.

Another problem arises when comparing the correlations of raw cognitive speed variables (in this case IT) with IQ versus MA: Age-related variance within the raw cognitive speed variable is likely to correlate with the age-related variance which is contained within MA variables (over and above the influence of ability variance). In contrast, age variance will have been

[1] [LIMITS correlates -.17 with CA, -.21 with RPM IQ and -.19 with PMA IQ. MCS1 correlates -.19 with CA, -.09 with RPM IQ ans -.15 with PMA IQ.]

removed from IQ scores via standardisation. Thus, the greater correlation of raw IT with MA compared to IQ may simply be a statistical epiphenomenon. A better indication of the true relationship between IQ and cognitive speed will be provided by standardising the speed variable within age groups. Any difference in size between the correlation of IQ and age-standardised IT, compared to the correlation of MA with raw IT (normalised) will thus reflect the influence of CA only.

Range-matched MA and IQ were calculated for two general ability variables: Raven's Matrices and total score on the PMA tests (VM+NF+SR+FG+WG+PS).[1] Comparisons of MA scores within the two age groups confirm a clear separation of the samples on both scales (for RPM MA: t=19.01 p<.0001; for PMA MA: t=22.21, p<.0001, one tailed).

IT (MCS2) was also normalised to the same distribution in the sample as a whole (equivalent to an MA score for IT, hereafter referred to as normalised raw IT, or <u>NRIT</u>) and within individual age groups (equivalent to an IQ score for IT - i.e. standardised IT, or <u>SIT</u>).

Shown in table 9.11 are the correlations of both estimates of <u>IQ</u> and <u>MA</u> (RPM & PMA) with raw IT, normalised raw IT <u>(NRIT)</u> and standardised IT <u>(SIT)</u>. For the purposes of comparison, the correlation of IT with chronological age (CA) is also shown, although it must be noted that this variable is not normally distributed, as the others are. Boxes in table 9.11 indicate comparable variables. For example, correlations between SIT and IQ are highlighted because both variables are age standardized normalized scores having the same range. Similarly, correlations between NRIT and MA are highlighted since, although both are normalised to the same range, they have not been age-standardized and hence remain remain influenced by (or share) age-related variance.

[1] [The normalised distributions of both MA variables range between z-scores -2.3 and +2.3, with a mean of 0 and a standard deviation of 1.]

Table 9.11

Correlations of MA and CA with raw IT scores, normalised raw IT scores (NRIT) and IT standardised within age groups (SIT).
(Correlations between comparable variables are shown in boxes)

RAW IT = Unadjusted IT
NRIT = Raw IT scores normalised within the whole sample (range = Z-2.3 to+2.3)
SIT = Raw IT scores standardised *within age groups* via normalisation (range as for NRIT)

	LIMITS			MCS1			*MCS2		
	Raw IT	NRIT	SIT	Raw IT	NRIT	SIT	Raw IT	NRIT	SIT
RPMIQ	-.21	-.20	$\boxed{-.20}$	-.09	-.14	$\boxed{-.14}$	-.37	-.38	$\boxed{-.38}$
RPMMA	-.29	$\boxed{-.28}$	-.19	-.19	$\boxed{-.23}$	-.13	-.39	$\boxed{-.41}$	-.33
PMAIQ	-.19	-.18	$\boxed{-.21}$	-.15	-.21	$\boxed{-.20}$	-.24	-.25	$\boxed{-.24}$
PMAMA	-.25	$\boxed{-.25}$	-.13	-.20	$\boxed{-.24}$	-.10	-.27	$\boxed{-.26}$	-.16
CA	-.17	-.20	0.00	-.19	-.19	0.00	-.14	-.14	0.00

As can be seen from table 9.11, normalising IT variables makes little difference to the size of their correlations with IQ and MA. This is not surprising in view of the normality of the raw variables.

For all three IT measures the correlation of normalised raw IT with MA is slightly higher than the correlation between standardised IT and IQ. The differences are small, however, indicating the very small amount of covariation between CA and IT. Where IT is measured by the most reliable method (MCS2) there is a negligible difference between the size of the MA-NRIT and IQ-IT correlations, confirming that CA has effectively no influence on IT.

CORRELATIONS BETWEEN INSPECTION TIME AND IQ IN SAMPLES DIFFERING IN ABILITY LEVEL

To investigate the hypothesis that IT-IQ correlations will be higher at lower IQ levels, the sample was split at the mean (SIQ), using either of two criteria, to produce - in each case- two groups of equal range, containing approximately equal numbers of subjects. The alternative split criteria were Raven's Matrices IQ and global PMA IQ (both re-standardised within age groups, to the same distribution). Raven's Matrices is the individual test with the highest g-loading in both age-groups. Global PMA IQ is calculated by standardising the total score on all the tests within the PMA. Selection on the basis of either criterion is therefore likely to produce a valid division of subjects in terms of general ability. In both comparisons, the 'below-average' and 'above average' sub-samples were selected so as to have the ranges -2.3 to 0 and 0 to +2.3, respectively, on age-standardised IQ (SIQ).

Table 9.12 Mean, standard deviation and range of RPM SIQ in samples above and below the overall sample average, and in both groups combined.
(Figures are expressed in standard score units.)

	N	Mean	S.D.	Range
Below Average IQ	60	-.75	.59	2.21 (-2.30 to -.09)
Above Average IQ	55	.82	.57	2.21 (+.09 to +2.30)
Whole sample	115	.00	1.00	4.60 (-2.3 to +2.3)

As can be seen from table 9.12, both sub-groups have identical ranges and approximately equal standard deviations on age-standardised RPM IQ. Mean SIQs in the two groups are approximately equal distances from the overall sample mean of 0. (The mean z scores of -.75 and .82 are equivalent to IQ scores of 89 and 112, relative to a full sample mean of 100.)

Correlations between IT and RPM IQ in 'low' and 'high' IQ groups.

Correlations between RPM IQ and IT in the two sub-groups, and in the whole sample, are contrasted in table 9.13. Age variance has been removed from the IT estimates by standardising within age-groups. Shown in parentheses are the relevant correlations after applying Pearson's correction for the effects of direct range-restriction.

Table 9.13
Correlations between Inspection Time (SIT) and Raven's Matrices IQ, in samples drawn from the lower and upper halves of the IQ Distribution

	LIMITS	MCS1	MCS2	Composite IT
Below Average RPM IQ	.07 (.12)	-.12 (-.20)	-.18 (-.30)	-.05 (-.11)
Above Average RPM IQ	-.28** (-.46)	.04 (.07)	-.19 (-.32)	-.21* (-.34)
Whole Sample	-.20*	-.14	-.38****	-.31***

[*p<.05, **p<.025, ***p<.005, one tailed]
(Figures in parentheses have been corrected for range restriction.)

As expected, correlations between IQ and IT are generally higher in the full sample than in either of the range-restricted groups. The exception to this is the correlation between RPM IQ and LIMITS in the above-average sub-group, which is higher than is the case in the full sample. This is explained by the low positive correlation between RPM IQ and LIMITS in the below-average sample. The correlation of RPM IQ with MCS1 approaches zero in the above-average sub-group, compared to a correlation of -.12 in the below-average sub-group. Using MCS2; the most reliable of the individual estimates; the correlation between IT and IQ is almost identical in the below-average and above-average IQ samples. For the composite IT variable, a near-zero correlation in the below-average sample contrasts with a correlation of .21 in the sample having higher IQ. Correcting for range-restriction increases the size of all correlations. In line with Frearson et al.'s (1988) findings, 'correction' boosts the size of several correlations to a level above the reference value for the full sample.

Correlations between RPM IQ and IT, in the samples below and above mean RPM IQ, were compared using Fisher's z. Only the uncorrected correlations are suitable for comparison in this way since, as pointed out by Ree et al (1994), the sampling distribution of corrected correlations is unknown. Results are shown below:

Table 9.14 Comparing correlations of RPM IQ (SIQ) and each estimate of IT (SIT) in samples of below and above-average RPM IQ, using Fisher's z test.

	LIMITS	MCS1	MCS2	Comp IT
Below vs Above Average IQ	1.53	-.78	-.06	-.85

[Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

In no case is the correlation between IQ and IT significantly stronger in samples drawn from the lower, versus the upper, half of the IQ distribution. There is thus no support for hypothesis 6 where IQ is estimated using Raven's Matrices. The difference between samples' IT-IQ correlations decreases with subsequent testings, falling to near zero where IT is estimated by MCS2, which - it has been argued - is the most reliable and valid method.

Comparison of samples below and above mean PMA IQ.

Dividing the sample at the mean on the global PMA scale, produces two groups of equal range on this IQ variable, as shown in table 9.15.

Table 9.15 Global PMA IQ (SIQ) and IT (SIT) in samples below and above the IQ mean.

	N	Mean	S.D.	Range
Below Average	59	-.75	.58	2.3 (-2.3 to 0)
Above Average	58	.80	.60	2.3 (0 to +2.3)
Whole sample	115	.00	1.00	4.6 (-2.3 to +2.3)

Correlations between IT and PMA IQ in the resultant sub-groups are as follows:

Table 9.16 Correlations between Inspection Time and Global PMA IQ, in samples drawn from the lower and upper halves of the IQ Distribution.(Shown in brackets is the correlation after correction for range restriction)

	LIMITS	MCS1	MCS2	Comp IT
Below Average PMA IQ	.19 (.35)	-.05 (-.10)	-.04 (-.09)	.07 (-.12)
Above Average PMA IQ	-.29 (-.41)	-.16 (-.26)	-.15 (-.24)	-.27 (-.38)
Whole Sample	-.18	-.21	-.26	-.27

[Critical values of r for a one-tailed test .21 ($p<.05$), .25 ($p<.025$), .32 ($p<.005$)

As when the sample is split by RPM IQ, the correlation between PMA IQ and IT is lower in the range restricted groups than in the whole sample, where IT is estimated using the method of constant stimuli. In the case of LIMITS, however, IT and IQ are more highly negatively related in the above-average sample than in the sample as a whole, since the correlation among subjects of below average IQ is in the opposite direction.*
By all three individual estimates, and using the composite IT variable, the negative correlation between IQ and IT is stronger in the sample having higher IQ, a result contrary to that hypothesed.

*(This could be a case of sign reversal after range restriction, as described by Ree et al, 1994, but the explanation offered seems more likely.)

For IT LIMITS and for the Composite IT variable, the correlations with IQ are positive in the sample with lower IQ. Comparing the size of correlations in the two groups using Fisher's z, the difference is not significant for any IT estimate. [LIMITS z= -.56, MCS1 z =-.59; MCS2 z= -.59; Composite IT z=-1.09.]* Once again, there is no support or the hypothesis that the relationship between IT and IQ will be more strongly negative in samples of lower versus higher IQ.

*The critical values of z, for significance at the five and ten percent levels (one-tailed) are 1.65 and 1.29. respectively. For LIMITS and Comp IT the sign of r for the below average group was reversed before computing z, so as not to yield a spuriously high estimate of the difference between the low and high IQ groups.

THE RELATIONSHIP BETWEEN IT AND IQ IN SAMPLES DIFFERING IN INSPECTION TIME.

The hypothesis that it is mental speed which underlies psychometric differentiation is central to the theories of Anderson, Spearman and others. To test this hypothesis, the combined sample (N=115) was split into two groups according each of the three inspection time estimates (LIMITS, MCS1, MCS2) and the composite IT variable derived from the average of the three (Comp IT). This yielded four 'fast' and four 'slow' groups. In each case, the sample was divided at the mean of the distribution of normalised raw IT, ensuring that the 'fast' and 'slow' groups had equal ranges. Within each sub-sample each IT variable was correlated with psychometric IQ (SIQ). The relevant coefficients were then corrected for the effects of explicit range restriction on the IT variable.

Splitting the sample according to IT estimated using the method of limits, produces subgroups with the following characteristics:

Table 9.17 Mean, standard devation and range of IT LIMITS (normalised raw IT) in 'slow' and 'fast' sub-samples formed by splitting the total sample at the mean.

	N	Mean	S.D.	Range
Above Average (short) IT	57	-.78	.57	2.3 (-2.3 to 0)
Below Average (long) IT	58	.80	.60	2.3 (0 to +2.3)
Whole sample	115	.00	1.00	4.6 (-2.3 to +2.3)

Inspection time (LIMITS) and eleven estimates of IQ (10 individual tests plus global PMA) were correlated in each of these two subsamples. The results of this analysis are shown in table 9.18

Table 9.18 Correlations between IT LIMITS (NRIT) and IQ (SIQ) in subsamples with below and above average inspection time estimated by the method of LIMITS. [Correlation coefficients shown in parentheses have been corrected for range restriction.]

	FASTER THAN AVERAGE		SLOWER THAN AVERAGE		ALL SUBJECTS	Difference between rIT-IQ in fast vs slow samples (Fisher's z)*
Raven's Matrices	-.01	(-.02)	-.30	(-.43)	-.19	1.56
Spatial Relations	-.08	(-.14)	-.13	(-.21)	-.25	.26
Figure Grouping	-.02	(-.04)	-.10	(-.16)	-.05	.42
Perceptual Speed	-.10	(-.17)	-.23	(-.35)	-.22	.70
Verbal Meaning	-.03	(-.05)	-.14	(-.23)	-.06	.58
Word Grouping	.02	(.04)	.00	(.01)	.05	.08
Word Fluency	.08	(.14)	.00	(.01)	-.01	.39
Figures of Speech	-.11	(-.19)	-.16	(-.25)	-.15	.27
Number Facility	-.05	(-.09)	-.13	(-.21)	-.09	.42
Digit Span	-.17	(-.28)	-.24	(-.36)	-.09	.38
{Global PMA IQ	-.05	(-.09)	-.21	(-.32)	-.18	1.47 }

*[Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

As shown in the above table, where the sample is split into a 'slow' and a 'fast' group according to the LIMITS estimate of IT, the size of the IT-IQ correlation is greater in the slower sample in eight out of ten cases, as predicted by hypothesis 7. The differences are not statistically significant at the .05 level, however, according to Fisher's z. For the most g-loaded of the tests (Raven's Matrices) and using global PMA IQ as an estimate of general intelligence, the difference in the strength of the IT-IQ correlation, is moderately significant (p<.10 in both cases).

Next, the sample was divided according to the second estimate of Inspection Time (MCS1). Sub-sample characteristics are shown in table 9.19.

Table 9.19 Mean, standard devation and range of IT MCS1 (normalised raw IT) in 'slow' and 'fast' subsamples formed by splitting the total sample at the mean.

	N	Mean	S.D.	Range
Above Average (short) IT	58	-.77	.58	2.3 (-2.3 to 0)
Below Average (long) IT	59	.78	.61	2.3 (0 to +2.3)
Whole sample	115	.00	1.00	4.6 (-2.3 to +2.3)

Correlations between IT (MCS1) and IQ in these 'fast' and 'slow' groups are given in table 9.20.

Table 9.20 Correlations between IT MCS1 (NRIT) and IQ (SIQ) in subsamples with below and above average inspection time estimated by the first method of constant stimuli.[Correlation coefficients shown in parentheses have been corrected for range restriction.]

	FASTER THAN AVERAGE		SLOWER THAN AVERAGE		ALL SUBJECTS	Difference between rIT-IQ in fast vs slow samples (Fisher's z)*
Raven's Matrices	-.07	(.12)	.12	(.20)	-.14	-1.00
Spatial Relations	-.05	(-.09)	-.05	(-.08)	-.21	.00
Figure Grouping	.12	(.21)	.22	(.37)	-.02	-.54
Perceptual Speed	-.06	(-.10)	.00	(.00)	-.14	-.33
Verbal Meaning	-.32	(-.58)	.12	(.20)	-.19	-2.38
Word Grouping	-.02	(-.03)	.17	(.28)	.01	-1.01
Word Fluency	.03	(.05)	-.15	(-.25)	-.15	.95
Figures of Speech	-.01	(-.02)	-.01	(-.02)	-.15	.00
Number Facility	-.02	(-.03)	.14	(.23)	-.03	-.85
Digit Span	-.03	(-.05)	.09	(.15)	-.09	-.63
{Global PMA IQ	-.22	(-.36)	.11	(.18)	-.21	1.05}

[*Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

Where the sample is split by MCS1, the least reliable of the three IT estimates, there are a greater number of negative IT-IQ correlations in the 'fast' than in the 'slow' group. The groups differ significantly in the size of the IT-IQ correlation in only one out of the ten cases (r IT & Verbal Meaning) and this occurs in the opposite direction to that predicted.

Splitting the sample according to IT estimated using the method of constant stimuli, second estimate, yields the following sub-groups:

Table 9.21 Mean, standard devation and range of IT MCS2 (normalised raw IT) in 'slow' and 'fast' subsamples formed by splitting the total sample at the mean.

	N	Mean	S.D.	Range
Above Average (short) IT	58	-.77	.57	2.3 (-2.3 to 0)
Below Average (long) IT	59	.78	.60	2.3 (0 to +2.3)
Whole sample	115	.00	1.00	4.6 (-2.3 to +2.3)

Correlations between IT (MCS2) and IQ in the fast and slow samples selected by MCS2 are shown below:

Table 9.22 Correlations between IT MCS2 (NRIT) and psychometric IQ (SIQ) in subsamples with below and above average inspection time estimated by the second method of constant stimuli.

	FASTER THAN AVERAGE		SLOWER THAN AVERAGE		ALL SUBJECTS	Difference between rIT-IQ in fast vs slow samples (Fisher's z)*
Raven's Matrices	-.26	(-.43)	-.38	(-.56)	-.38	.71
Spatial Relations	-.17	(-.29)	-.42	(-.61)	-.38	1.45
Figure Grouping	-.07	(-.12)	-.38	(-.56)	-.28	1.74
Perceptual Speed	-.19	(-.32)	-.25	(-.40)	-.13	.33
Verbal Meaning	.02	(.04)	.03	(.05)	-.07	-.05
Word Grouping	-.01	(-.02)	-.07	(-.12)	.00	.32
Word Fluency	.07	(.12)	-.31	(-.48)	-.08	2.06
Figures of Speech	-.08	(-.14)	-.07	(-.12)	-.13	-.05
Number Facility	-.05	(-.09)	-.23	(-.37)	-.11	.97
Digit Span	.07	(.12)	-.01	(-.02)	-.09	.42
{Global PMA IQ	-.16	(-.27)	-.33	(-.50)	-.25	.96 }

[*Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

As table 9.22 shows, when the sample is split into 'slow' and 'fast' subgroups according to IT estimated by the second method of constant stimuli there is a clear tendency for the negative association between IT and IQ to be stronger in the 'slow' group. This is evident in ten out of the eleven comparisons shown in table 9.22. The difference is statistically significant in only two cases (FG, WF), nevertheless, eight out of the eleven comparisons shown in table 9.22 are in the predicted direction. Furthermore, the correlations of IT with IQ are higher in the slower sample using both estimates of general intelligence (RPM & PMA). It has previously been argued that, of the three IT estimates collected in this study, MCS2 is the most valid and reliable, followed by LIMITS, with MCS1 rather unreliable.

Next, the sample was divided at the mean on the composite IT variable, derived by averaging the three raw estimates and fitting the product to a normal curve.

Table 9.23 Mean, standard devation and range of Composite IT in 'slow' and 'fast' subsamples formed by splitting the total sample at the mean.

	N	Mean	S.D.	Range
Above Average (short) IT	58	-.77	.58	2.3 (-2.3 to 0)
Below Average (long) IT	58	.80	.60	2.3 (0 to +2.3)
Whole sample	115	.00	1.00	4.6 (-2.3 to +2.3)

Correlations between composite IT and IQ in these 'fast' and 'slow' subgroups are shown in table 9.24.

Table 9.24 Correlations between IT and psychometric IQ (SIQ) in subsamples with below and above average inspection time, estimated using the composite variable.

	FASTER THAN AVERAGE		SLOWER THAN AVERAGE		ALL SUBJECTS	Difference between rIT-IQ in fast vs slow samples (Fisher's z)*
Raven's Matrices	-.06	(-.08)	-.21	(-.36)	-.29	.82
Spatial Relations	-.10	(-.16)	-.30	(-.43)	-.36	1.13
Figure Grouping	-.09	(-.15)	-.12	(-.21)	-.12	.16
Perceptual Speed	-.10	(-.16)	-.27	(-.44)	-.24	.95
Verbal Meaning	.10	(.16)	-.09	(-.15)	-.11	-.05
Word Grouping	-.08	(-.14)	-.16	(-.23)	.06	.44
Word Fluency	.08	(.14)	-.24	(-.36)	-.06	1.75
Figures of Speech	-.15	(-.24)	-.22	(-.37)	-.17	.39
Number Facility	-.13	(-.21)	-.20	(-.33)	-.09	.39
Digit Span	.11	(.19)	-.03	(-.05)	-.11	-.43
{Global PMA IQ	-.21	(-.34)	-.26	(-.42)	-.26	.29 }

[*Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

As can be seen table 9.24, when the sample is split according to average/composite IT, negative correlations between IT and IQ are consistently stronger in the group with slower-than-average inspection times. Although the difference is only significant in one of the eleven comparisons reported in table 9.24 (Word Fluency) the data suggest some support for hypothesis 7.

To summarise: There is a tendency for the association between Inspection Time and IQ to be stronger in samples with longer-than-average IT, compared to those with shorter-than average IT. This tendency is particularly marked where the sample is split according to the most reliable of the three estimates (MCS2) and by the composite IT variable derived from the average of the three estimates. Although the results do not have the statistical significance to allow for firm conclusions to be drawn, they do provide modest support for the theory that basic speed of information processing is a predictor of cognitive differentiation.

Chapter 10

Results for Study 4:

Testing Anderson's Model Using Shape Discrimination Speed as an Indicator of the Efficiency of the Spatial-Analogical Processor (SP2)

Response times in the samples of nine and twelve year olds

The average time taken to discriminate between identical and mirror image shapes, at each of the four levels of orientation disparity, is shown in table 10.1. Separate figures are given for pairs of identical and pairs of mirror image shapes, and for the average of 'same' and 'mirror image' pairs at each orientation. Mean response time across all conditions has also been calculated for each age group. For each variable, the age difference has been assessed using an unrelated t-test (one-tailed).

Table 10.1 Mean and Standard Deviation of Shape Discrimination Time at Four Levels of Orientation Disparity in Nine and Twelve Year Olds Samples.
(In all cases, RT was recorded in centiseconds.)

SHAPES	ORIENTATION DISPARITY	NINE YR OLDS MEAN (S.D.)	TWELVE YR OLDS MEAN (S.D.)	AGE DIFFERENCE t value	P=
SAME	0	262.68 (73.04)	193.02 (55.16)	5.78	.0001
SAME	60	382.99 (111.53)	286.28 (79.11)	5.37	.0001
SAME	120	406.42 (109.22)	314.36 (96.02)	4.75	.0001
SAME	180	458.13 (123.89)	363.37 (121.08)	4.15	.0001
DIFFERENT	0	300.74 (79.66)	208.94 (51.03)	7.37	.0001
DIFFERENT	60	449.79 (124.46)	332.69 (91.86)	5.75	.0001
DIFFERENT	120	479.92 (130.12)	372.00 (113.50)	4.74	.0001
DIFFERENT	180	454.33 (113.99)	345.06 (104.01)	5.37	.0001
BOTH TYPES	0	281.71 (64.38)	200.98 (48.59)	7.60	.0001
BOTH TYPES	60	416.39 (105.20)	309.48 (79.67)	6.15	.0001
BOTH TYPES	120	442.67 (107.70)	343.18 (96.69)	5.21	.0001
BOTH TYPES	180	456.23 (105.28)	354.21 (106.92)	5.15	.0001
OVERALL MEAN RT		399.25 (86.21)	301.96 (77.86)	6.35	.0001

As can be seen from table 10.1, the expected increase in RT with increasing shape orientation disparity is evident in both age groups. There is a slight deviation from this pattern when mirror image pairs are considered separately, with response times at 180 degrees being slightly faster than those at 120 degrees. At orientation disparities of 0, 60 and 120 degrees, RT is longer for mirror image than for identical shape pairs, although for mirror image pairs RT is slightly faster at 180 degrees than at 120 degrees.

Twelve year olds have significantly faster response times than nine year olds, at every level of orientation disparity. Averaged across all conditions, the response times of twelve year olds are approximately 10 seconds faster than those of the younger sample. Hypothesis 4 is therefore supported.

Mental rotation speed (as opposed to rotation speed + discrimination time) can be estimated by subtracting the time taken to discriminate between shapes at the same orientation (0 degrees) from that taken to respond when one of the shapes is rotated by 180 degrees, relative to the other. Using the combined results for same and mirror image pairs, this difference is 174.52 centiseconds (s.d. 82.39) for nine year olds, and 153.23 (80.79) for twelve year olds. Although the older sample are marginally faster, the difference is not significant at the 5 percent level, using an unrelated t-test ($t=1.40$ $p=.17$ two-tailed, or .08 one-tailed). [1]

[1] This finding does not conflict with the hypothesis that age groups differ in RT, since the variables thought to underlie age differences in RT mainly affect the response element of the task, which is removed from this measure of shape rotation speed.

Correlations Between Shape Discrimination Times and Psychometric Test Scores in Nine and Twelve Year Old Samples.

Shown in tables 10.2 a and b are the correlations between psychometric test scores and shape discrimination times within each age group. Once again, results for 'same' and 'mirror image' stimuli are shown separately and combined. Average RT across conditions is also included, along with the estimate of shape rotation speed (mean RT at 180-0 degrees).

It will already have been noted, from table 10.1, that for several of these parameters, the standard deviation of RT is marginally smaller in the older sample. To eliminate the effects of differential s.d. all of the RT variables shown have been normalised to the same distribution, as have the psychometric test variables. In all cases the range is from -2.3 to +2.3 Z scores, with a mean of 0 and a standard deviation of 1.0.

Tables 10.2 a & b. Correlations between psychometric test scores and parameters of the shape discrimination task, where both types of variable have been normalised to the same distribution.

10.2 (a) NINE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
SAME 0	.06	-.10	-.18	-.16	-.03	.06	-.26	.00	.12	-.01
SAME 60	-.20	-.15	-.26	-.37	-.07	-.08	-.26	-.11	-.04	-.17
SAME 120	-.03	-.06	-.01	-.33	-.04	.06	-.11	-.01	.19	-.12
SAME 180	-.05	.10	-.07	-.32	-.04	.19	-.15	-.04	.11	-.11
DIFFERENT 0	-.20	-.07	-.21	-.24	-.28	-.06	-.32	.08	-.04	-.19
DIFFERENT 60	-.03	-.10	-.06	-.32	-.11	-.04	-.24	-.18	-.01	-.18
DIFFERENT 120	.01	.08	.05	-.22	.00	.11	-.18	-.10	.10	-.20
DIFFERENT 180	.05	-.07	-.01	-.21	-.04	.06	-.06	-.09	-.00	-.08
ALL 0	-.11	-.13	-.26	-.23	-.18	-.01	-.35	.00	-.00	-.09
ALL 60	-.14	-.15	-.20	-.36	-.09	-.05	-.27	-.18	-.04	-.20
ALL 120	.00	.06	.02	-.29	-.01	.12	-.18	-.08	.16	-.21
ALL 180	-.03	.02	-.06	-.32	-.08	.09	-.10	-.04	.05	-.14
MEAN RT	-.10	-.06	-.14	-.36	-.10	.06	-.22	-.09	.03	-.18
[ALL 180-0	.09	.13	.15	-.19	.06	.18	.19	-.05	.06	-.02]

10.2 (b) TWELVE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
SAME 0	-.45	-.19	-.23	-.44	-.30	-.14	-.25	-.25	-.07	-.05
SAME 60	-.51	-.25	-.30	-.57	-.41	-.00	-.23	-.16	-.04	-.00
SAME 120	-.49	-.28	-.18	-.45	-.25	-.16	-.15	-.25	-.09	-.12
SAME 180	-.48	-.30	-.35	-.54	-.40	-.17	-.18	-.14	-.10	-.04
DIFFERENT 0	-.52	-.22	-.24	-.60	-.42	-.05	-.28	-.22	-.06	-.07
DIFFERENT 60	-.44	-.29	-.40	-.57	-.40	-.16	-.20	-.20	-.19	-.05
DIFFERENT 120	-.41	-.30	-.31	-.50	-.35	-.12	-.10	-.16	-.17	-.06
DIFFERENT 180	-.49	-.30	-.22	-.40	-.25	-.16	-.11	-.18	-.02	-.08
ALL 0	-.47	-.19	-.20	-.52	-.38	-.10	-.25	-.20	.01	-.02
ALL 60	-.49	-.31	-.40	-.58	-.42	-.10	-.23	-.18	-.12	-.01
ALL 120	-.48	-.31	-.26	-.49	-.32	-.15	-.12	-.20	-.15	-.07
ALL 180	-.48	-.29	-.29	-.49	-.35	-.13	-.14	-.14	-.05	-.03
MEAN RT	-.51	-.33	-.35	-.54	-.38	-.15	-.19	-.19	-.11	-.05
[ALL 180-0	-.37	-.30	-.24	-.35	-.23	-.13	-.01	-.06	-.04	-.03]

In both age groups, the correlations between RT and IQ variables are in the expected, negative, direction (hypothesis 1).

In both samples, the test having the greatest degree of association with shape discrimination time is *Spatial Relations*. This is unsurprising in view of the fact that the stimuli in this test are very similar to those in the RT task. Across shape stimuli, the correlation of RT with *Spatial Relations* test score is -.36 for nine year olds and -.54 for twelve year olds. After SR, the test which correlates to the greatest extent with RT is, in nine year olds, *Perceptual Speed* (-.22) and in twelve year olds *Raven's Matrices* (-.52). Correlations with shape discrimination speed are also high in the case of *Figure Grouping*, although this is only the case for twelve year olds. In general, correlations between shape RT and IQ are appreciably higher in the older of the two samples.[1]

[1] When correlations between mean RT and psychometric variables are averaged, the figure is -.013 for nine year olds [ignoring differences in sign] and -.28 for twelve year olds.

The supposed estimate of shape rotation speed (RT at 180 degrees of orientation disparity minus RT at 0 degrees) correlates negatively with all ten psychometric variables in the sample of twelve year olds. In contrast, the corresponding correlations in the sample of nine year olds are low and not all in the same direction. In both samples average shape discrimination time is more strongly predictive of psychometric test scores than is rotation time, estimated using the subtractive method. For this reason, the decision was taken to focus attention on response times in later analyses.

Reliability of the Shape Discrimination Task.

Although age differences in test variances have been controlled for in the above matrices, it is important to consider whether the higher correlations of RT with IQ in twelve year olds are a by-product of greater test reliability in that sample. The reliability of the shape discrimination task can be estimated from the correlations between response times for identical and mirror image pairs at each level of orientation disparity (parallel forms reliability) and from the overall correlation between response times to same and mirror image stimuli across all conditions. These figures are shown below for the two age groups.

Table 10.3 Correlations between discrimination times for identical and mirror image shapes at each level of orientation disparity, and across all four conditions.*

	NINE	TWELVE
0 DEGREES	.45	.73
60 DEGREES	.62	.74
120 DEGREES	.60	.73
180 DEGREES	.64	.85
ALL CONDITIONS	.76	.90

*The correlations shown are between normalised variables which correlate with raw RT variables at above .98 within age groups. For raw RT, the overall correlation between RT to same and mirror image shapes is .75 for nine year olds and .92 for twelve year olds.]

It can be seen from table 10.3 that response times to same and mirror image pairs are more strongly associated in the older of the two age groups, indicating a potentially important contribution to the higher correlations between IQ and RT in that sample.

Reliability coefficients for each of the ten psychometric test variables, within each age group, were reported in an earlier section. These revealed only minor differences between in the two samples and it was found that correction for the effects of unreliability did little to alter the age differences (or lack thereof) in the sizes of inter-subtest correlations. Nonetheless, reliability estimates for both contributing variables are required in the formula for computing corrected correlations. Tables 10.4 a and b. show the correlations between psychometric test scores and RT parameters of the shape discrimination task after correction for unreliability.

Tables 10.4 a & b. Correlations between psychometric variables and average shape discrimination times at each level of orientation disparity and across all conditions. Correlations coefficients have been corrected for measurement unreliability.

10.4 (a) NINE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
ALL 0	-.18	-.21	-.41	-.39	-.32	-.02	-.54	.01	-.00	-.16
ALL 60	-.19	-.20	-.27	-.52	-.14	-.08	-.36	-.23	-.05	-.29
ALL 120	.00	.08	.03	-.42	-.02	.18	-.25	-.10	.24	-.31
ALL 180	-.04	.03	-.08	-.46	-.12	.14	-.13	-.04	.07	-.19
MEAN RT	-.13	-.07	-.17	-.47	-.14	.09	-.27	-.10	.04	-.23
[Before correction	-.10	-.06	-.14	-.36	-.10	.06	-.22	-.09	.03	-.18]

10.4 (b) TWELVE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
ALL 0	-.61	-.25	-.25	-.71	-.65	-.15	-.31	-.24	.01	-.03
ALL 60	-.63	-.39	-.48	-.79	-.72	-.16	-.28	-.22	-.15	-.01
ALL 120	-.62	-.40	-.32	-.67	-.54	-.23	-.15	-.24	-.20	-.08
ALL 180	-.58	-.34	-.33	-.61	-.54	-.19	-.16	-.16	-.06	-.04
MEAN RT	-.59	-.38	-.38	-.66	-.59	-.21	-.21	-.21	-.12	-.06
[Before correction	-.51	-.33	-.35	-.54	-.38	-.15	-.19	-.19	-.11	-.05]

As would be expected, correction for unreliability increases the size of all correlations, with the difference most pronounced in the sample for whom RT reliability is lowest (i.e. nine year olds). The tendency for RT-IQ correlations to be higher in twelve, than in nine year olds remains, however, although the difference is reduced somewhat. The average of the correlations between mean RT and each of the ten psychometric variables is, -0.17 in the nine year old sample and -.34 in twelve year olds The corresponding averages before correction are -.01 and -.28, respectively.

To test the significance of the difference between corresponding RT-IQ correlations in the two age groups, all coefficients were transformed to values of Fisher's Z and compared using the z test. Significant differences are underlined in table 10.5.

Table 10.5 Values of z indicating the degree to which the correlations between shape discrimination time and psychometric test scores differ in the samples of 9 versus 12 year olds

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
0	<u>3.08</u>	.19	-.95	<u>3.47</u>	<u>2.32</u>	.72	.25	1.31	-.05	-.69
60	<u>3.12</u>	1.09	1.33	<u>3.61</u>	<u>3.93</u>	.42	.53	.56	.51	-1.51
120	<u>3.81</u>	<u>2.62</u>	<u>1.87</u>	<u>2.65</u>	<u>3.09</u>	<u>2.15</u>	.13	1.01	<u>2.32</u>	-1.22
180	<u>3.27</u>	<u>1.99</u>	1.33	<u>2.00</u>	<u>2.56</u>	1.72	.47	.72	.65	-.81
MEAN RT	<u>3.03</u>	<u>1.70</u>	1.21	<u>2.19</u>	<u>2.80</u>	1.56	.42	.85	.87	-.93

[Critical Value of z for significance at 5% level =1.65 10%=1.29, one-tailed]

Positive values in table 10.5 indicate higher RT-IQ correlations in the older sample. As was the case before correcting for unreliability, most RT-IQ correlations are higher in the older sample, although the reverse is true in the case of correlations between RT and score on the *Figures of Speech* test. The age differences in the size of correlations between average RT and psychometric ability are statistically significant in the case of only four of the ten IQ tests, however: *Raven's Matrices*, *Figure Grouping, Spatial Relations* and *Verbal Meaning*. Nevertheless, pronounced age differences are also found in the case of *Number Facility* and *Word Grouping*, although they are not significant at the five percent level, using a one tailed test. These results do not clearly support hypothesis 5.

[It is interesting to note that the age difference is most marked for the three tests requiring the greatest degree of spatial ability, but not for the test requiring only simple visual scanning and identification, i.e. *Perceptual Speed.*]

Relationship between the g-loadings of psychometric tests and the size of their correlations with shape discrimination speed.

To test the hypothesis that the size of the RT-IQ correlation is related to the g-loading of the IQ variable in question, average response time, across all conditions, was correlated with the g-loading of each test variable, as described in an earlier section. Resulting Spearman correlations in the samples of nine and twelve year olds were low, but in the predicted direction. For 9 and 12 year olds, respectively, Rho = -.006 (ns) and -.264 (p<.01, one tailed). In the full sample, there was a strong correlation between tests' g-loadings and the size of their correlations with average shape discrimination time (Rho= -.61, p<.005, one-tailed), supporting hypothesis 2.

CORRELATIONS OF SHAPE DISCRIMINATION TIME WITH IQ VERSUS MENTAL AGE.

In the combined samples (N=115) average shape discrimination time correlates more highly with MA (-.51 in the case of RPM MA and -.53 in the case of PMA MA) than with IQ (-.26 for RPM IQ and -.29 for PMA IQ).
As was discussed when reporting results for the Inspection Time task, this difference reflects the influence of chronological age, which elevates the MA-RT correlation (since both are affected by it) and suppresses the correlation between IQ and RT (since IQ *is* age-standardised whilst RT is not). The correlation between average shape discrimination time and chronological age is -.53.
The influence of CA on MA-RT correlations can be removed using partial correlations and the influence of CA on IQ-RT correlations can be removed by standardising RT within age groups. Results, following these

adjustments, are shown in tables 10.6 a and b. In both cases, the CA-contaminated correlation is shown in brackets, for the purposes of comparison.

Table 10.6 (a) Correlation between MA and mean shape RT when the effect of CA has been partialled out [and before this is done]

	Average Response Time, Normalised in the whole sample.
Raven's Matrices MA	-.32 (-.51)
Primary Mental Abilities MA	-.28 (-.53)

Table 10.6(b) Correlation between IQ and mean RT, where RT has also been standardised within age groups (and where RT has not been age-standardised)

	Average Response Time, Standardised, via normalisation, within each age group.
Raven's Matrices IQ	-.31 (-.26)
Primary Mental Abilities IQ	-.32 (-.29)

The difference between IT-MA and IT-IQ correlations is eliminated when CA is controlled for. These figures are interesting since they reveal the true association between ability and speed of response. Although not as marked as the association between CA and RT (r = -.53) the association between ability and RT is clear (-.31 for RT and RPM IQ, -.32 for RT and PMA IQ).

Relationship between IQ and Speed of Shape Discrimination at Different Ability Levels.

Hypothesis 6 predicts that RT-IQ correlations will be stronger at lower IQ levels. Following the same procedure used when considering *Inspection Time* data, the sample was split at the mean (SIQ), using either of two criteria, to produce - in each case- two groups of equal range, containing approximately equal numbers of subjects. The alternative split criteria were Raven's Matrices IQ and global PMA IQ (both re-standardised within age groups, to the same distribution). Tables 10.7 a & b show the correlations between each IQ variable and RT in sub-samples from the lower and upper halves of the IQ distribution. Each correlation has been corrected twice, once for the effects of range restriction and again for the effects of measurement unreliability. For the latter purpose, the reliability of each variable was re-calculated within each subgroup.

Table 10.7 (a) Correlations of Ravens Matrices (RPM) IQ with Average Shape RT (age-standardised) where the sample has been selected on the basis of RPM IQ.

LOW (RPM Z= +2.3 to 0) [N=60]			HIGH (RPM 0 to -2.3) [N=55]			All Subjects [N=115]	
Uncorrected	Range-Corrected	Corrected for Unreliability	Uncorrected	Range-Corrected	Corrected for Unreliability	Uncorrected	Corrected For reliability
.04	(.07)	[.08]	-.01	(-.02)	[-.02]	-.31	[-.35]

Table 10.7 (b) Correlations of Primary Mental Abilities (PMA) IQ with Average Shape RT (age-standardised) where the sample has been selected on the basis of PMA IQ.

LOW (PMA Z= +2.3 to 0) [N=59]			HIGH (PMA 0 to -2.3) [N=58]			All Subjects [N=115]	
Uncorrected	Range-Corrected	Corrected for Unreliability	Uncorrected	Range-Corrected	Corrected for Unreliability	Uncorrected	Corrected for reliability
-.27	(-.43)	[-.50]	-.16	(-.26)	[-.28]	-.33	[-.36]

According to table 10.7 (b) the correlation between mean shape discrimination time and global Primary Mental Abilities IQ is stronger in a sample consisting of subjects scoring at the 50th percentile, or below, on this variable (r = -.50, after correction) than in a sample drawn from the upper half of the IQ distribution (r = -.28). The difference is moderately significant, according to Fisher's z test ($z = 1.37$, $p < .10$, one-tailed) offering some support for hypothesis 6. The results are less clear, however, when the sample is divided according to Raven's Matrices IQ (table 10.7 a). In this case, restriction of the IQ range severely attenuates the correlation between RT and IQ in both sub-groups. Corrections for the effects of direct range restriction and test unreliabilities do little to change this, making it impossible to draw firm conclusions about ability group differences in the strength of the relationship between Shape RT and IQ.

These mixed results do not allow for any firm conclusions to be drawn concerning the association between shape discrimination RT and IQ at different levels of the ability spectrum.

Relationship between IQ and Speed of Shape Discrimination in Samples with 'Fast' versus 'Slow' Response Times.

Following the procedure adopted with the Inspection Time results, the sample was split into a 'fast' and a 'slow' group on the basis of average shape discrimination time (standard normal scores). Correlations between IQ and RT were then recalculated. The resulting figures are shown below. Given in parentheses are the relevant correlation coefficients after correction for the effects of direct range restriction and test unreliability.

Table 10.8
Correlations between average shape discrimination time (Shape RT) and two estimates of general intelligence in samples with Below (Fast) and Above (Slow) Average Shape RT.
[UC=Uncorrected, Rg Cr =Corrected for range restriction, Rel Cr = Corrected for unreliability

	'FAST' (Z=-2.3 to 0)			'SLOW' (Z=0 to +2.3)			ALL SUBJECTS (Z=-2.3 to +2.3)	
	UC	Rg Cr	Rel Cr	UC	Rg Cr	Rel Cr	UC	Rel Cr
RPM IQ	-.35	(-.54)	[-.58]	-.26	(-.30)	[-.35]	-.31	[-.35]
PMA IQ	-.27	(-.44)	[-.47]	-.11	(-.18)	[-.20]	-.32	[-.35]

The negative correlation between average shape discrimination time and IQ is markedly stronger in the 'faster' of the two samples. This result is in the opposite direction to that predicted by the differentiation hypothesis. When correlations between IQ and RT are compared in the fast and slow groups using Fisher's z test none of the differences are significant at the 5 percent level, using a two-tailed test. Hypothesis 8 is not supported by these findings.

Correlations between Shape RT and IQ in groups differing in Inspection Time

Next, the sample was split into a 'fast' and a 'slow' group on the basis of Inspection Time (MCS2).

Table10.9 Correlations between average shape discrimination time (Shape RT) and two estimates of general intelligence in samples with Below (Fast) and Above (Slow) Average IT [UC=Uncorrected, Rg Cr =Corrected for range restriction, Rel Cr = Corrected for Unreliability]

	'IT FAST' (Z=-2.3 to 0)			'IT SLOW' (Z=0 to +2.3)		
	UC	Rg Cr	Rel Cr	UC	Rg Cr	Rel Cr
RPM	-.32	-.54	-.60	-.25	-.42	-.49
PMA	-.38	-.59	-.64	-.21	-.35	-.40

As when the sample was divided on the basis of RT, correlations between shape RT and IQ are higher in subjects with shorter-than-average Inspection Time than in those with longer-than-average IT. This result runs contrary to the hypothesis 7.

Comparing the correlations between RPM IQ and Shape RT in samples with slow versus fast IT using Fisher's z test, the difference is not significant at the 5% level, using a one-tailed test ($z=.83$). The difference is significant, however, in the case of PMA-RT correlations ($z=1.77$, $p<.05$).

Taken together, these findings do not support the hypothesis that correlations between Reaction Time and IQ will be stronger in samples with slower mental speed.

Chapter 11

Results for Study 5: Testing Anderson's Model Using Letter Identification Speed as an Indicator of the Efficiency of the Verbal Propositional Processor (SP1)

Comparing nine and twelve year old samples

Table 11.1 shows the average time taken to identify pairs of letters as being physically or phonetically the 'same' or 'different', in each age group. Results are given separately for each letter combination and also for the average RT within each condition (name & physical identification). The average difference between name identification time and physical identification time (theoretically indicating speed of access to phonological codes) is also shown. For all variables, age differences have been tested for significance using unrelated t-tests (one-tailed).

Table 11.1 Time taken to make a 'same' vs 'different' discrimination between pairs of letters where the criterion of discrimination is physical or phonetic (name) identity (ms).

STIMULI	CONDITION	NINE YEAR OLDS MEAN	(SD)	TWELVE YEAR OLDS MEAN	(SD)	AGE DIFFERENCE t value	P=
AA, aa	Physical ID	1039.33	(143.54)	776.65	(119.71)	10.67	.0001
AA, aa	Name ID	1079.31	(157.59)	882.93	(171.80)	6.39	.0001
Aa, aA	Physical ID	1131.75	(214.55)	836.84	(153.06)	8.50	.0001
Aa, aA	Name ID	1389.39	(217.09)	1066.00	(188.29)	8.54	.0001
AB, ab	Physical ID	1116.84	(211.55)	839.52	(171.90)	7.72	.0001
AB, ab	Name ID	1353.14	(262.38)	1024.24	(211.18)	7.41	.0001
Ab, bA	Physical ID	1095.68	(213.72)	814.09	(145.90)	8.26	.0001
Ab, bA	Name ID	1354.11	(253.77)	1065.26	(223.42)	6.48	.0001
MEAN RT FOR PHYSICAL ID		1095.90	(174.48)	816.78	(134.63)	9.62	.0001
MEAN RT FOR NAME ID		1293.86	(196.40)	1009.61	(182.79)	8.04	.0001
OVERALL AVERAGE RT		1194.95	(159.62)	913.19	(145.93)	9.88	.0001
DIFFERENCE N-P		198.84	(189.78)	192.83	(133.74)	0.20	.422 (ns)

It can be seen from table 11.1 that, for all stimulus conditions and for the averaged physical and name identification conditions, twelve year olds respond significantly faster than do nine year olds. In all cases the difference is significant to the .0001 level, using a one-tailed t-test. This result supports hypothesis 4.

For all letter pair combinations, response times are greater when the required discrimination is based on phonetic identity than when it is based on physical identity. With respect to the difference between these two conditions (hypothetically indicating the speed of access to lexical codes, following physical identification), the two age groups do not differ significantly (t=.20, p=.422, one-tailed).

Correlations between IQ and Letter Discrimination Time in Nine Versus Twelve Year Olds.

As with the shape discrimination task, the standard deviation of RT is slightly lower in the twelve year old sample. For this reason, all variables were normalised to the same distribution (Z =-2.3 to +2.3) before comparing correlations in the two age groups.

Tables 11.2 a & b show the correlations between psychometric variables and letter discrimination times in the two age groups. Figures are given separately for *physical* and *name* discrimination times, for the average RT across conditions, and for the difference between *name* and *physical* identification times. Non-verbal, verbal and number tasks are clustered separately, for convenience. (Corresponding figures for the combined sample can be found in appendix 11.1.)

Tables 11.2 a & b. Correlations between IQ Variables and Letter Discrimination Time in 'Physical Identity' and 'Name Identity' Conditions, between IQ and mean RT across conditions, and between IQ and Speed of lexical access as measured by the subtraction of P from N.

11.2 (a) Nine Year Olds.

	Physical ID	Name ID	Average RT	N - P
Raven's Matrices	-.01	-.16	-.10	-.16
Spatial Relations	.05	-.10	-.04	-.18
Figure Grouping	-.17	-.24	-.25	-.16
Perceptual Speed	-.29	-.34	-.38	-.07
Verbal Meaning	-.12	-.17	-.18	-.11
Word Grouping	-.13	-.23	-.23	-.15
Word Fluency	.10	-.08	-.01	-.22
Figures of Speech	-.01	-.24	-.16	-.26
Number Facility	-.12	-.28	-.25	-.14
Digit Span	.09	-.04	.05	-.11
Average	-.06	-.19	-.16	-.16

11.2 (b) Twelve Year Olds.

	Physical ID	Name ID	Average RT	N - P
Raven's Matrices	-.40	-.42	-.45	-.17
Spatial Relations	-.30	-.26	-.30	-.14
Figure Grouping	-.21	-.16	-.22	-.05
Perceptual Speed	-.27	-.41	-.39	-.28
Verbal Meaning	.03	-.04	-.01	-.16
Word Grouping	.03	-.13	-.10	-.21
Word Fluency	-.24	-.39	-.35	-.28
Figures of Speech	-.06	-.24	-.17	-.28
Number Facility	-.45	-.36	-.44	-.05
Digit Span	.04	.02	.02	-.01
Average	-.18	-.24	-.24	-.16

[Critical values of *r*, using a one-tailed test, are .21 ($p<.05$), .25 ($p<.0225$), .325 ($p<.005$)]

In both age groups, letter discrimination time (in both the *physical* and *name* identification conditions) and speed of lexical access (N-P) correlate negatively with the majority of IQ variables, as predicted by hypothesis 1. On average, negative correlations between letter identification time and IQ are stronger in the *name* identity condition than in the *physical* identity condition.[1]

[1] This presumably reflects the greater cognitive load of the the latter task.

Before drawing conclusions about age differences in the strength of these correlations, it is important to consider the effects of age differences in the reliability of the letter identification task.

Reliability of the Letter Identification Task in the Two Age Groups.

Correlations between response times to the three categories of letter pair in the 'physical' and 'name' identity conditions are shown, for each age group in tables 11.3 a and b. Also shown is the average of these inter-correlations within each condition (i.e. physical or name identification) and between averaged 'physical' and averaged 'name' discrimination times.

Table 11.3 (a) Correlations between letter discrimination times (normalised) within and between conditions, in the nine year old sample.

PHYSICAL IDENTIFICATION	AA,aa	Aa, aA	AB, ab	NAME IDENTIFICATION	AA,aa	Aa, aA	AB, ab
Aa, aA	.639			Aa, aA	.638		
AB, ab	.775	.664		AB, ab	.635	.699	
Ab, bA	.709	.794	.740	Ab, bA	.673	.738	.753

AVERAGE of correlations between 3 letter pair types in physical identity condition = .720
AVERAGE of correlations between 3 letter pair types in name identity condition =.689

Correlation between averaged 'physical' identification time and averaged 'name' identification time = .521

Table 11.3 (b) Correlations between letter discrimination times (normalised)within and between conditions, in the twelve year old sample.

PHYSICAL IDENTIFICATION	AA,aa	Aa, aA	AB, ab	NAME IDENTIFICATION	AA,aa	Aa, aA	AB, ab
Aa, aA	.790			Aa, aA	.835		
AB, ab	.734	.792		AB, ab	.789	.827	
Ab, bA	.747	.783	.787	Ab, bA	.746	.810	.770

AVERAGE of correlations between 3 letter pair types in physical identity condition = .772
AVERAGE of correlations between 3 letter pair types in name identity condition =.796

Correlation between averaged 'physical' identification time and averaged 'name' identification times = .684

Averaging the intercorrelations between letter discrimination times for each type of letter pair, within conditions, gives an estimate of the internal reliability of the *physical* and *name* identification tasks. Estimated in this way, reliability is slightly higher in the older sample: .77 in twelve year olds, compared to .72 in nine year olds. For name identification the corresponding figures are .80 and .69 (rounded to two decimal places).
The correlation between average RT in the *physical identity* condition and average RT in the *name identity* conditions is also higher in the older sample (.68, compared to .52 in nine year olds).

Since the slightly lower reliability of letter discrimination times in the younger sample may contribute to age differences in the size of RT-IQ correlations, all correlations were corrected for attenuation. Where the to-be-corrected correlation is between IQ and *either* average Physical RT *or* Name RT, corrections were based on the within-condition estimates of reliability (i.e. physical: age nine = .72, twelve =.77; name: age nine = .69, twelve = .80). In the case of correlations between IQ and the difference between RT in Name and Physical identification conditions (N-P) and between IQ and the overall average RT across the two conditions, the estimate of reliability is the correlation between average RT in the 'physical' condition and average RT in the 'name' condition.(i.e. .52 for nine year olds and .68 for twelve year olds).

Tables 11.4 a) and b) show the corrected correlations between IQ variables and letter discrimination time in the 'Physical Identity' and 'Name Identity' conditions, between IQ and mean RT across conditions, and between IQ and speed of lexical access (as measured by the subtraction of P from N).

Tables 11.4 a & b. Correlations between IQ and Parameters of the Letter Discrimination Task in Nine and Twelve Year Olds, After Correcting for Test Unreliability.

11.4 (a) Nine Year Olds.

	Physical ID	Name ID	Average RT	N-P
RPM	-.02	-.21	-.15	-.23
SR	.06	-.14	-.06	-.28
FG	-.23	-.35	-.41	-.27
PS	-.36	-.43	-.56	-.10
VM	-.15	-.22	-.27	-.16
WG	-.19	-.33	-.39	-.25
WF	.12	-.09	-.02	-.32
FS	-.01	-.32	-.25	-.40
NF	-.14	-.35	-.37	-.21
DS	.12	-.06	.08	-.17
Average	-.08	-.25	-.24	-.24

11.4 (b) Twelve Year Olds

	Physical ID	Name ID	Average RT	N-P
RPM	-.50	-.52	-.60	-.23
SR	-.39	-.34	-.42	-.20
FG	-.34	-.26	-.38	-.08
PS	-.32	-.47	-.48	-.35
VM	.04	-.05	-.01	-.21
WG	.05	-.19	-.16	-.34
WF	-.29	-.46	-.43	-.35
FS	-.07	-.29	-.22	-.36
NF	-.52	-.41	-.55	-.06
DS	.05	.03	.03	-.02
Average	-.23	-.30	-.32	-.22

[Critical values of *r*, using a one-tailed test, are .21 ($p<.05$), .25 ($p<.0225$), .325 ($p<.005$)]

In the sample of nine year olds, the average (negative) correlation between *physical* identification time and IQ variables remains low (-.08) after correcting for unreliability, although the correlation with *Perceptual Speed* exceeds .35. As in the uncorrected data, the averaged correlations between *name* RT and IQ and between *overall average* RT and IQ are much higher (-.25 and -.24, respectively). The IQ variables correlating most strongly with average letter RT are *Perceptual Speed*

(-.56), *Figure Grouping* (-.41), *Word Grouping* (-.39) and *Number Facility* (-.35). The average correlation of IQ with speed of lexical access (N-P) is also relatively high (-.24), with this variable correlating most strongly with the two tests of ideational fluency *Figures of Speech* (-.40) and *Word Fluency* (-.32)

In twelve year olds, as in the younger sample, the average correlation between RT and IQ is stronger in the case of *name* identification time (-.30) than *physical* identification time (-.23), with the average correlation between IQ and *overall* mean RT being slightly greater (-.32). The average correlation between *speed of lexical access* (N-P) and IQ (r=-.22) is similar to that between IQ and *physical* RT. In twelve year olds, the non-verbal (RPM, SR, FG, PS) and Number Facility variables correlate to a greater degree with *physical* RT than do verbal IQ variables. This pattern is also evident for *name* identification time, although Word Fluency correlates at -.46 with this variable. In the case of *lexical access speed* (N-P), correlations are much stronger with verbal than with non-verbal IQ variables.

To reveal the overall pattern of age differences in IQ-RT correlations, and to test the significance of any differences, the corrected RT-IQ correlations obtained by the two samples were compared using Fisher's z. The results of this analysis are shown in table 11.5.

Table 11.5 Comparison of IQ-RT correlations in nine and twelve year old samples using Fisher's z, after correction for measurement unreliability.

	Physical ID	Name ID	Average RT	N - P
RPM	2.78	1.89	2.83	-.20
SR	2.50	1.15	2.04	-.28
FG	.60	-.47	-.20	-.39
PS	-.19	.27	-.52	.08
VM	-1.02	-.92	-1.35	-.14
WG	-1.25	-.80	-1.31	-.09
WF	2.17	2.08	2.30	-.15
FS	.29	-.19	-.11	-.24
NF	2.24	.35	1.18	-.34
DS	.38	-.44	.27	-.34

[Critical Value of Z for significance at 5% level =1.65 10%=1.29, one-tailed]

Positive values in table 11.5 indicate higher correlations in the older of the two samples. Where average *'physical'* identification time is correlated with IQ scores, seven out of ten correlations are higher in the older sample. When average *'name'* identification time or *overall average* RT are correlated with IQ, the number of correlations which are higher in the older sample is equal to the number which are lower (5 &5). Age differences in the size of correlations between average RT and IQ are statistically significant in only three out of ten cases. [1]

None of the correlations between IQ and purported speed of lexical access (N-P) differ significantly in the two age groups.[2]

Taken together, these results offer support for hypothesis 5.

Relationship Between the g-Loadings of Psychometric Tests and the Size of their Correlations with Letter Identification Speed.

Following the procedure used with the Inspection Time and Shape Discrimination tasks, Spearman correlations were computed between the g-loading of each psychometric test variable and its correlation with average Letter Discrimination Speed. The value of Rho was, for nine year olds -.56 (p=.09) and for twelve year olds, -.59 (p=.07). In the full sample, the association is even stronger (Rho= .66, p<.05). This result supports the hypothesis that the strength of the (negative) correlation between a psychometric test variable and reaction time can be predicted from that variable's loading on the general factor (hypothesis 2).

[1] Correlations of physical identification time with IQ are significantly higher in twelve year olds in the case of Raven's Matrices, Spatial Relations,Word Fluency and Number Facility. For name identification time, correlations are significantly higher in the older sample only for Raven's Matrices and Word Fluency. For overall average RT correlations with IQ are significantly higher in twelve year olds in the case of Raven's Matrices, Spatial Relations and Verbal Fluency.

[2] Although, in general, correlations are marginally *lower* in the older group.

Correlations of Letter Identification Time with IQ versus MA.

Shown in table 11.6 is the correlation of average letter discrimination time with CA, and with IQ and MA according to the Raven's Matrices and Primary Mental Abilities tests. Shown in brackets are the correlations between RT and IQ after the influence of age variance has been removed via standardisation of the RT variable and the correlations between MA and RT with the influence of CA partialled out.

Table 11.6 Correlation of average letter identification time with Chronological Age and with MA and IQ in the Raven's Matrices and PMA Tests.

r CA & RT: -.68

RPM MA & RT:	-.53 (-.27)	RPM IQ & RT	-.19 (-.28)
PMA MA & RT	-.69 (-.39)	PMA IQ & RT	-.29 (-.38)

Response time in the letter discrimination task is more strongly predicted by (chronological) age than by IQ.

The Relationship Between IQ and Letter Discrimination Time at Different Levels of General Ability.

As with the results for the Inspection Time and Shape discrimination tasks, the sample was split at the mean according to either of two estimates of general intelligence - Raven's Matrices IQ or PMA IQ. To remove the influence of chronological age, and to ensure that each data split produced two groups of equal variance, all variables were normalised within age groups, before calculating correlations between RT parameters and psychometric test scores.

Table 11.7 shows the correlations of RPM and PMA IQ with average letter discrimination time, in samples scoring above and below average on the relevant IQ variable. Also shown, for comparison, are the relevant correlations after correcting for the effects of range-restriction and test unreliability within sub-groups.

Tables 11.7 Correlations of IQ with Average Response Time in the Letter Discrimination Task, in samples drawn from the lower and upper halves of the IQ distribution.

Split Criterion	Category	Correlation with Average Letter Discrimination Time: Uncorrected	Range-Corrected	Corrected for range and unreliability	Difference (Fisher's z)
RPM	'LOW'	-.05	-.09	-.16	2.21 ($p<.05$)
RPM	'HIGH'	-.29	-.43	-.61	
PMA	'LOW'	-.29	-.43	-.59	1.30 ($p<.10$)
PMA	'HIGH'	-.19	-.32	-.41	

The above results reveal an inconsistent relationship between general ability level and the strength of the correlation between letter identification time and IQ. When the sample is split according to Raven's Matrices IQ, the negative correlation between IQ and average letter discrimination time is markedly stronger in the above-average group. In contrast, sample division by PMA IQ results in nonsignificantly higher IQ -RT correlations in the group with below average IQ.

Relationship between IQ and Letter Discrimination Speed in Samples with 'Fast' versus 'Slow' Response Times.

To test the hypothesis that the strength of RT-IQ correlations will be greater in groups with slower, as compared to faster, response times, the sample was split into two groups according to average letter identification time. RT was then correlated with two estimates of general intelligence (RPM IQ & PMA IQ) in each sub-sample.

The relevant correlations are shown in table 11.8.

Table 11.8 Correlation between average Letter identification time, across conditions, and IQ in groups with below-average (FAST) and above-average (SLOW) RT on this variable. (UC=uncorrected Rg C = Corrected for range-restriction Rel C= Corrected for unreliability)

	'FAST'(Z= -2.3 to 0)			'SLOW' (Z= 0 to +2.3)		
	UC	Rg C	Rel C	UC	Rg C	Rel C
RPM IQ	-.28	-.45	-.66	-.04	-.07	-.10
PMA IQ	-.25	-.41	-.57	-.03	-.05	-.06

When the sample is divided into a 'fast' and 'slow' group on the basis of average letter discrimination time, the negative relationship between RT and IQ is markedly stronger in the group with shorter response times. This result is in the opposite direction to that predicted by hypothesis 8.

Correlations between Letter RT and IQ in groups differing in Inspection Time

Next, the sample was divided at the mean according to the most reliable estimate of Inspection Time (MCS2) and RT-IQ correlations recalculated. Results are shown in table 11.9.

Table 11.9 Correlation between average Letter Identification time, across conditions, and IQ in groups with below-average (FAST) and above-average (SLOW) INSPECTION TIME. (UC=uncorrected Rg C = Corrected for range-restriction Rel C= Corrected for unreliability)

	'FAST IT' (Z= -2.3 to 0)			'SLOW IT' (Z= 0 to +2.3)		
	UC	Rg C	Rel C	UC	Rg C	Rel C
RPM IQ	-.37	-.56	-.76	-. 23	-.38	-.50
PMA IQ	-.59	-.41	-.78	-. 19	-. 31	-. 41

As was found when the sample was divided according to letter identification time, division by Inspection Time results in higher RT-IQ correlations in the sample with *faster* IT. This result is contrary to hypothesis 7.

<u>The Association Between Verbal Processing Speed and Spatial-Processing Speed in Samples Differing in IQ and Inspection Time.</u>

The final hypothesis to be addressed predicts that *"RT to verbal-propositional stimuli will correlate less with RT to spatial-analogical stimuli at high levels of IQ or general speed".* (Hypothesis 9.)

To test this hypothesis, the sample was divided at the mean according to *Inspection Time* using the estimate argued earlier to be the most reliable (MCS2) or either of the two measures of general intelligence (*Raven's Matrices IQ* and *Primary Mental Abilities IQ).* Average response time in the <u>Letter Discrimination Task</u> was correlated with average RT in the <u>Shape Discrimination Task</u> in the two sub-samples formed by each selection method. Correlations were then corrected for test unreliability, using the reliability coefficients derived within each sub-group. The resultant correlation coefficients are shown below.

<u>Table 11.10</u>

<u>Correlations between letter identification time and shape discrimination time in sub-samples drawn from the lower and upper halves of the distribution of Inspection Time or IQ. *</u>

Selection criterion	Category [range]	Uncorrected		Corrected for unreliability	
		r	difference z	r	difference z
IT (MCS2)	'FAST' [Z=-2.3 to 0]	.38	.30(n.s.)	.52	.47 (n.s.)
IT (MCS2)	'SLOW' [Z=0 to +2.3]	.42		.61	
IQ (RPM)	'HIGH' [Z=0 to +2.3]	.30	.32 (n.s)	.41	.94 (n.s.)
IQ (RPM)	'LOW' [Z=-2.3 to 0]	.35		.54	
IQ (PMA)	'HIGH' [Z=0 to +2.3]	.32	.27(n.s.)	.43	1.00 (n.s.)
IQ (PMA)	'LOW' [Z=-2.3 to 0]	.37		.57	

*[All variables are standardised within age groups, via normalisation.]

It can be seen from table 11.10 that letter identification time and shape discrimination time are more strongly correlated in the samples having longer inspection times and lower IQs than in the faster, brighter samples. The difference in correlation sizes becomes more pronounced after correction for sub-group differences in the reliability of the measures. Although this finding is in line with the hypotheses, inter sub-group differences in correlation size are not statistically significant, according to Fisher's z test. This result may be interpreted as offering partial support for hypotheses 6 and 8.

Chapter 12.

Summary and Discussion of Results for Studies 3, 4 and 5

Since studies 3, 4 and 5 were all designed to address Anderson's theory, it is appropriate to consider them together. Indeed, chapters 9-11 would have been merged, had the results not been so lengthy. In view of this, the results will be reviewed and examined with respect to the hypotheses arising from Anderson's model, as detailed in chapter 7.

Hypothesis 1

The first hypothesis raised in chapter 7, predicts that *individual differences in inspection time (and reaction time) will be negatively correlated with performance on psychometric tests.* Whilst this hypothesis has been confirmed in many previous experiments (see e.g. Eysenck, 1982; Vernon, 1987; Deary, 1996) it is important to check that the data arising from this study are consistent with this general finding, to establish the validity of the measures.

The results confirm that *Inspection Time*, *Shape Discrimination Time* and *Letter Discrimination Time* correlate negatively with all or most psychometric test variables within the samples of nine and twelve year olds and in the combined samples. Average correlations in the combined sample are shown below:

Table 12.1 Correlation of general intelligence with Inspection Time and two measures of response time (shape discrimination/letter discrimination) in the full sample (N=115)

	Composite IT	Mean Shape RT	Mean Letter RT
Raven's Matrices IQ	-.31	-.31	-.28
Average PMA IQ	-.27	-.32	-.38

(All variables have been standardised within age to eliminate the influence of CA on the correlation. All coefficients are significant at the 0.05 level, or better, using a one-tailed test.)

These findings are consistent with those of previous investigators and lend support to the theory that mental speed underlies performance on complex cognitive (or IQ) tests. All but one of the coefficients shown in table 12.1 exceed the magic 0.3 suggested by Hunt (1980) to represent a theoretically significant association between an elementary cognitive task and IQ. These figures have not been corrected for measurement unreliability, a process which raises their values, however, their size is likely to have been magnified by the fact that most of the variables represent composite measures.

Hypothesis 2

The second prediction arising from Anderson's model, is that *the degree to which psychometric test scores correlate with measures of cognitive speed will be associated with the ability tests' loadings on the general factor (g).*

The results offer partial support for this hypothesis. Considering the sample as a whole: The correlation between *Inspection Time* and IQ is predicted by the g-loading of the IQ variable, where the most reliable and valid estimate of IT (MCS2) is used (*Rho* = -.43 $p<.05$). The relationship is non-significant for the first estimate (LIMITS) and for the composite IT measure (*Rho* = -.16 in both cases), although it is in the predicted direction. For the least reliable measure (MCS1) the relationship is in the opposite direction (*Rho* = .47 $p<.05$). There is also a significant association between IQ tests' loadings on the first principal component and their correlations with average *Shape Discrimination Time* (*Rho* = -.61, $p<.05$). An even stronger relationship between tests' g-loadings and the size of their correlations with RT is evident in the case of average *Letter Identification Time* (*Rho*= -.66 ($p<.05$).

These results are consistent with the theory that mental speed (the *e* of Anderson's Basic Processing Mechanism) underlies Spearman's g.

Hypothesis 3

The third hypothesis addresses Anderson's unique view that 'intelligence does not develop'. He theorises that inspection time is a relatively pure measure of mental speed and that mental speed does not change during childhood. This leads to the prediction that samples aged nine and twelve years, matched for IQ, *will not* differ significantly with respect to average Inspection Time.

Using the first (LIMITS) and second (MCS1) estimates, and the average of the three, twelve year olds have significantly shorter inspection times than nine year olds ($t=1.82$, 2.09 and 2.59 respectively; $p<.05$ in all three cases). The age difference in inspection time declines over the testing sequence, however, suggesting that it may be due to age-related error which lessens with practice. Where IT is estimated using the final and most reliable measure (MCS2) there is no significant difference between the two age groups ($t=1.51$). The evidence, in regard to hypothesis 3 is, therefore, inconclusive.

Hypothesis 4

Anderson's contention that response time measures are more prone to the effects of learned strategies leads to the fourth hypothesis that *average RT will differ significantly between age groups.*

Supporting this, is the finding that twelve year olds have significantly faster response times than nine year olds on both the Shape Discrimination and Letter Identification tasks. Average shape discrimination time is 3.02 seconds in twelve year olds and 3.99 seconds in nine year olds ($t=6.35$, $p<.0001$). The corresponding figures for average Letter Identification time are 9.13 and 11.95 sec., respectively ($t = 9.88$, $p<.0001$).

Hypothesis 5

Anderson theorises that the strength of Spearman's g is primarily dependent upon mental speed. His hypothesis that this is unchanging during child development also leads to the prediction that *there will be no difference between age groups in the size of correlations between IQ and measures of cognitive processing speed.*

None of the forty correlations between *Inspection Time* and IQ (ten IQ tests and three individual + one composite IT estimate) are significantly different in the two age groups. Furthermore, the age difference in IT-IQ correlation size is smallest in the case of MCS2, argued to be the most reliable of the three individual estimates. Correlations between average *Shape Discrimination Time* and IQ, are significantly different in the two age groups for only three of the ten IQ tests (Raven's Matrices, Verbal Meaning, Spatial Relations and Figure Grouping). In these three cases, as in five of the nonsignificant comparisons, correlation sizes are higher in the older group. Similarly, correlations between average *Letter Identification Time* and IQ are significantly different in the two age groups in only three cases (Raven's Matrices, Spatial Relations and Word Fluency). Whilst in the latter the relationship between letter RT and IQ is stronger for the twelve year olds, there is no general tendency for the correlations to be higher or lower in either sample.

In short, there is no consistent tendency for IQ measures to correlate with Inspection Time or Response Time measures to a greater or lesser extent in samples aged nine and twelve. This finding offers some support for Anderson's theory that the differentiation effect is not associated with age differences. As described in chapter 2, this goes against the findings and theories of other researchers, such as Spearman (1925), Burt (1954) and Garrett (1946).

Hypothesis 6

The sixth hypothesis stems from the theory that positive manifold declines at higher levels of general intelligence. *Measures of cognitive processing speed are thus expected to correlate more highly with psychometric test variables in samples composed of subjects with lower versus higher IQ.*

Dividing the sample at the mean for Raven's Matrices IQ and correlating IT and RT with RPM IQ in the 'below average' and 'above average' subgroups, yields inconsistent results:

There is no difference in the sizes of correlations between Inspection Time estimates and RPM IQ in the two samples, when the most reliable estimate of IT (MCS2) is used ('below average' r =-.30, 'above average' r= -.32, z=-.06 ns). Using LIMITS, the negative correlation is non-significantly stronger in the 'above average' sample, contrary to the hypothesis (-.46 vs .12, z=1.53). Using MCS1 the relationship is in the direction predicted by hypothesis 6, being stronger in the 'below average' group (r = -.20, compared to .07, z=-.78 ns), but the difference is not statistically significant. Composite IT , however, correlates more strongly with RPM IQ in the 'above average' sample (-.34 vs -.11 z=-.85 ns)

Correlations between RPM IQ and average *Shape Discrimination Time* do not differ significantly between samples with 'above average' and 'below average' RPM IQ. (In fact the RT-IQ correlations in both groups are very low -.02 and .08, respectively.)

The correlation between average *Letter Discrimination Time* and Raven's Matrices IQ is higher in the sample with 'above-average' RPM IQ compared to that for the 'below-average' sample, contrary to hypothesis 6 (r=-.61 and -.16, respectively, z=2.32, p<.05, one-tailed).

When the sample is divided according to average Primary Mental Abilities IQ, correlations between all three estimates of *Inspection Time* and this IQ variable are stronger in the 'above average' sample, contrary to hypothesis 6. (Comparing samples with lower and higher mean IQ: For LIMITS r= .35 vs -.41, z=-.56 ns; MCS1 r=-.10 vs -.26, z= -.59 ns; MCS2 r=-.09 vs -.24, z=-.59 ns; Comp IT r= -.12 vs -.38, z= -1.09 ns.)

In contrast, average *Shape Discrimination Time* and IQ are more strongly related in the sample scoring below-average on PMA IQ than in the above-average sample (r= .50 & .28, respectively, z=1.37). The tendency for RT-IQ correlations to be higher in a sample with below-average PMA IQ is also evident in the case of *Letter Discrimination Time.* (r= -.59 vs -.41, respectively, in the 'low' and 'high' IQ groups, z=1.30). Although neither of these differences are significant at the five percent level, using Fisher's z, they offer moderate support for the hypothesis (p<.10, in both cases, using a one-tailed test).

These mixed findings do not provide conclusive support for the hypothesis that mental speed and IQ are more strongly related in groups having lower, as compared to higher, IQ.

Hypothesis 7

A distinctive feature of Anderson's model of differentiation, is the emphasis he places on the role of individual differences in fundamental processing speed. This gives rise to the prediction that *measures of cognitive processing speed will correlate more highly with psychometric test variables in samples with slower-than average versus faster-than-average Inspection Time.*

Inspection Time and IQ were correlated in samples drawn from the lower and upper halves of the distribution of each IT variable. (The coefficients shown below have been corrected for range restriction.)

Where IT is estimated by the least reliable method (MCS1), there is no clear tendency for the size of IT-IQ correlations to be higher in the slower of the two groups: Correlations of MCS1 with Raven's Matrices IQ are -.12 and -.20, respectively, in the 'fast' and 'slow' groups (z=-1.00, ns). Correlations of MCS1 with global PMA IQ are -.36 and -.18, respectively, in the 'fast' and 'slow' groups (z=-1.05, ns).

Where IT is estimated by the method of LIMITS, correlations are higher in the slower sample for eight out of the ten IQ variables, in line with hypothesis 7. The difference is approaches significance for both estimates of general intelligence, Raven's Matrices IQ (r=-.43 vs -.02, respectively, z=1.56, p<.10) and PMA IQ (r= -.32 vs -.09, respectively, z=1.47 p<.10).

When the sample is divided according to the most reliable estimate of IT (MCS2), there is, again, a tendency for IT-IQ correlations to be higher in the sample with slower-than-average Inspection Time. The difference is significant in only two cases (IT with FG z=1.75, and WF z=2.06). The two estimates of general intelligence (RPM IQ and PMA IQ) correlate more highly with IT in the 'slow' than in the 'fast' sample, in line with the hypothesis, although the differences are not statistically significant according to Fisher's z test (RPM slow vs fast: -.56 vs -.43 z=.71, PMA slow vs fast: -.50 vs -.27, z=.96).

Taken together, these results provide only moderate support for the hypothesis that IT is a better predictor of IQ in samples with slower, as compared to faster, cognitive speed.

Correlations between two measures of general intelligence (Raven's Matrices IQ and Primary Mental Abilities IQ) and the two measures of response time (Shape Discrimination Time and Letter Identification Time) were also compared in the samples drawn from the lower and upper halves of the distribution of *Inspection Time*. Both measures of general intelligence correlate more highly with Shape Discrimination Time and with Letter Identification Time in the sample having *faster* mean Inspection Time (rRPM IQ & *Shape RT* - 'Fast' = -.58, 'Slow' = -.35; rPMA IQ & *Shape RT* 'Fast' = -.47, 'Slow' = -.20; rRPM IQ & *Letter RT* 'Fast' =-.76, ''Slow' = -.41). These findings reduce the strength of the evidence in support of hypothesis 7.

Hypothesis 8

Hypothesis 8 follows the same reasoning as hypothesis 7, predicting that *measures of cognitive processing speed will correlate more highly with psychometric test variables in groups with slower versus faster reaction times.*

Contrary to this hypothesis, average *Letter Discrimination Time* and general intelligence correlate more strongly in a sample having faster-than-average letter RT than in a sample with slower-than average RT. This is found both when general intelligence is estimated from Raven's Matrices IQ (r 'fast' = -.66, r 'slow' = -.10, corrected for attenuation) and Primary Mental Abilities IQ (r 'fast = -.57, r 'slow' = -.06, corrected). The differences in correlation sizes between 'fast' and 'slow' groups are not statistically significant according to Fisher's z.

The same pattern of results is found in the case of RT in the Shape Discrimination Task. Once again, correlations of PMA IQ and RPM IQ with average shape RT are higher in the sample drawn from the bottom half of the distribution of shape-RT (i.e. the 'faster' sample). The differences are not statistically significant (r SRT & RPM IQ:'fast'= -.58, 'slow' = -.35; r SRT & PMA IQ: 'fast -.47, 'slow' -.35).

There is no evidence to support hypothesis 8.

Hypothesis 9

In Anderson's model, the e-level of the BPM is most strongly predictive of the relationship between the two specific processors. This leads to the prediction that *speed of response to verbal-propositional stimuli* (hypothetically indicating the efficiency of SP1) *will correlate with RT to visuospatial stimuli* (SP2) *to a lesser degree in a sample of subjects having faster-than-average Inspection Time* (BPM speed) *than in a sample with slower-than average Inspection Time.*

To test this hypothesis, the sample was divided at the mean of the distribution of IT, measured using the most reliable method (MCS2). Average *Shape Discrimination Time* and average *Letter Identification Time* were correlated with each other in the 'Fast IT' and 'Slow IT' groups. The magnitude of the correlation between shape-RT and letter-RT was greater in the sample having 'slow' IT ('slow' r= .61, 'fast' r=.52), although the difference is not statistically significant (z=.47). This result is in line with hypothesis 9.

Assuming IT to be the central determinant of differences in general intelligence, an associated hypothesis is that *RT to Verbal and Spatial stimuli will be more strongly associated in samples of lower versus higher general intelligence.* To test this hypothesis, the sample was divided at the mean of either of the two general ability variables; Raven's Matrices IQ or PMA IQ; and RT-RT correlations computed for each subgroup. Correlations between Shape-RT and Letter-RT were higher in the sample with lower average IQ, using both split criteria. (High RPM IQ r=.30, Low RPM IQ r=.35 ; High PMA IQ r=.32, Low PMA IQ r=.37). In neither case, however, was the difference in the size of RT-RT correlations obtained in 'low' and 'high' IQ samples, statistically significant (RPM low vs high z=.94, PMA low vs high z=1.00). Nonetheless, this result offers further support for hypothesis 9.

CHAPTER 13

OVERVIEW AND CONCLUSIONS

The primary aims of this thesis have been to investigate the historical, theoretical and statistical underpinnings of the differentiation hypotheses, to test them empirically using traditional psychometric measures and to examine the features of cognitive architecture postulated by Anderson (e.g.1986) to underlie the reducing positive manifold effect. The main purpose of this chapter is to review the results of these investigations and to discuss their implications.

A historical review (chapters 1-3) revealed that theories of differentiation have appeared, intermittently, in the literature since Spearman discovered the law of positive manifold in the early years of this century. Studies which have examined the influence of age and, less commonly, ability on g have produced inconsistent results. Similarly, explanations for reported sample differences in the strength of g have varied widely, with different theorists emphasising inherent statistical bias, biological causes related to maturation or neurological speed, and differential learning experiences. It is undoubtedly due to these variations that interest in and knowledge about such hypotheses has ebbed and flowed over the course of the past seventy years. Whilst the 'age differentiation hypothesis', appears to have retained a tenuous foothold to the present day (e.g. Carroll, 1993), at the time research for this thesis began, the possibility that sample differences in general ability might influence the strength of Spearman's g had been overlooked by the majority of senior researchers in the field (e.g. Detterman and Daniel, 1989).

Anderson's model of minimal cognitive architecture (e.g. Anderson, 1986)

provided a modern explanation for the differentiation phenomenon and, at the outset of this research, had appeared unique. A fascinating outcome of the literature review, however, was the rediscovery of a very similar theory proposed by Charles Spearman as early as 1925. The relative neglect of the topic in the intervening years is noteworthy.
Both Anderson's and Spearman's theories hold that the main factor determining the extent of positive manifold is level of basic cognitive speed. The latter is thought also to predict level of general intelligence, hence sample differences in IQ will, according to this reasoning, predict the degree to which diverse abilities are correlated. Two main issues separate Anderson's theory from Spearman's. Firstly, Anderson maintains that mental speed (and consequently positive manifold) does not change over the course of child development. Spearman believes that speed increases during childhood, leading to the weakening of the g factor. The second main difference between the two theories is that in Anderson's model the primary effect of increased 'basic processing speed' is to release the capacities of two semi-independent 'specific processors' (SP1 and SP2) which give rise to spatial-analogical and verbal-propositional abilities, roughly representing the contralateral cognitive functions. These broad ability 'processors' are not separated in Spearman's model, where all specific abilities are treated similarly.

The studies reported in this thesis have not been concerned with examining the precise factorial structure of abilities in different groups. This may, as many theorists have speculated (e.g. Garrett,1946; Vernon,1965; Anastasi, 1970), be affected by specific learning experiences and hence vary across samples differing in age and ability. *The overriding aim of these studies has been to determine whether there are clear trends in the extent to which abilities are differentiated or integrated, depending on the age or IQ of the sample examined.* This is adequately covered within the theoretical models forwarded by Anderson and Spearman, and it is for this reason that these models (particularly the former) have been focused upon in the studies reported. Nonetheless, they encompass ideas expressed by many other theorists (e.g. Burt, 1954; Rienert, 1970).

Studies 1 and 2 examined the association between diverse psychometric tests at different ages and levels of ability. Studies 3, 4 and 5, were primarily concerned with testing Anderson's theories regarding the role of mental speed in predicting g, and with elucidating the mechanisms of his minimal cognitive architecture. The results of these studies are summarised in the following section.

Summary of results

Study 1. Examining developmental differences in positive manifold.

Study 1 examined the strength of Spearman's g in five age groups, ranging from eight to twelve years, who were tested using a battery of ten psychometric measures. Results based on the raw data suggested a clear negative relationship between increasing age and decreasing g-strength, in line with the age differentiation hypothesis of Burt (1954), Garrett (1946) and others. Closer examination of the data, however, revealed that test reliability and sample heterogeneity decreased systematically with increasing age. Whilst age differences in test reliability had a negligible influence on the apparent differentiation effect, range restriction effects were more obviously implicated. The issue of how to correct for these effects was examined in detail, noting the pros and cons of each method for dealing with this atypical case (explicit selection is a more common problem, whereas these range restriction effects were unintended and affected some variables more than others). Four methods were considered: reducing the ranges of the nine to twelve year olds to match

those of the eight year olds; correcting the correlations using Pearson's formulae; adjusting all score distributions to the same range via normalisation, and using partial correlations to remove the effect of age differences in range from the association between age and g-strength. Close examination of these methods revealed none to be ideal, although all but the first were attempted (the first was impractical since it drastically reduced the sample sizes of the oldest groups).
Improving the comparability of samples through normalisation did little to reduce the differentiation effect found in the raw data. It is possible, however, that ceiling effects in the older samples reduced the range of true scores in these groups, hence normalisation may not have fully ameliorated the problem of differential range. Algebraic correction for range restriction eliminated the differentiation effect, but it was acknowledged that this may have been due to the tendency of such formulae to overcompensate (Frearson et al., 1987). The slight U-shaped relationship between age and g-strength, observed after correction, lends some credence to this interpretation. Extracting the influence of age differences in variance by using partial correlations decreased, but preserved, the significant association between age and g-strength.

In short, the findings of study 1 were mixed and do not provide conclusive evidence for the hypothesis that g declines in strength during childhood. Neither do they fully support Anderson's prediction that age will not affect positive manifold since intelligence 'does not develop'.

Study 2: Ability-level and g.

Two comparisons were made to test the hypothesis that positive manifold will be weaker in samples of higher, versus lower, ability. The first of these (study 2a) compared the strength of Spearman's g in sub-groups drawn from the lower and upper halves of the distribution of each of ten psychometric test variables. No differences in g-strength were found, even after correction for differential test reliability. Closer examination of the data, however, revealed that whilst the mean IQs of the two sub-groups were approximately 88 and 112, respectively, on the variable used for selection, the means of the other nine (incidentally selected) variables tended to be much closer to one another, averaging 96 and 103 (a pattern many previous studies have failed to acknowledge). For this reason, it was decided to select three more widely separated sub-groups and perform the comparisons again (study 2b). The 'low', 'mid' and 'high' ability groups had average IQs on the selector variables, of 79, 100 and 121, respectively, and 94, 100 and 106 on the incidentally selected variables. Degree of positive manifold was again calculated for each sub-group, before and after correcting for differential test reliability. Results revealed no significant difference in the strength of g between the groups, contrary to the differentiation hypothesis. Since subjects of very low IQ were not studied, however, it was not possible to refute the hypothesis that Spearman's g is strongest within the very lowest ability ranges.

Studies 3, 4 and 5 attempted to test Anderson's hypotheses concerning the role of general mental speed in accounting for sample differences in positive manifold. In so doing, they extended the scope of the previous two studies by examining correlations among measures of cognitive processing speed and between the latter and IQ.

Study 3: Testing Anderson's model using inspection time as an indicator of the efficiency of the basic processing mechanism.

Study 3 compared samples of nine and twelve year olds and sub-groups differing in IQ level, with respect to their performance on *Inspection Time* measures and the relationship between IT and IQ, as predicted by

Anderson's differentiation hypothesis.

Results revealed a clear negative association between Inspection Time and IQ, replicating the findings of many previous investigators (see e.g. Vernon, 1987; Deary, 1996). The individual psychometric test with which the composite IT measure correlated most strongly in the full sample was Spatial Relations (a test of visuospatial manipulation ability), followed closely by Raven's Matrices (a non-verbal test of general intelligence). The reliabilities of the individual IT estimates were fairly low, although they increased over the testing sequence. The degree of association between IT and any of the ten IQ variables was predicted by the IQ test's loading on the general factor, in line with the hypothesis that speed underlies general intelligence, although this relationship was only significant for the most reliable of the three individual IT estimates (MCS2).

Comparing age groups:

Twelve year olds had significantly faster inspection times than did nine year olds, according to the first (LIMITS) and second (MCS1) estimates and for the composite IT measure, but not for the second and most reliable estimate (MCS2). Furthermore, the degree of difference between the two samples decreased with successive testings, implying a role for practice effects. (This was discussed in chapter 10). The evidence in favour of Anderson's hypothesis that inspection time does not differ between age groups, is therefore mixed.

The sizes of correlations between the IQ and IT variables were compared in the nine and twelve year old samples. In no case were these significantly different in the two age groups, a result which is compatible with Anderson's theory that mental speed underlies the differentiation phenomenon but is unchanging through development. The age difference in average IT-IQ correlation size was smallest in the case of the most reliable estimate.

Comparing samples differing in ability level:

The relative magnitude of IT-IQ correlations was compared in sub-groups of 'above average' and 'below average' ability, selected according to either of two estimates of general intelligence. Sample division by Raven's Matrices IQ yielded no consistent tendency for IT-IQ correlations to be either stronger or weaker in the two ability groups. For the most reliable estimate (MCS2) IT-IQ correlations were virtually identical in the two groups and there was little difference for MCS1. For LIMITS and the composite measure IT-IQ correlations were nonsignificantly higher in the 'above average' group, contrary to the hypothesis. When the sample was divided according to PMA IQ, the negative correlation between all four IT estimates (including composite IT) and PMA IQ was stronger in the sample of higher ability, although none of the differences were statistically significant. Study 3 thus yielded no evidence to support the hypothesis that correlations between IT and IQ will be stronger in samples of lower, compared to higher, general intelligence.

Comparing groups differing in inspection time:

To test Anderson's theory that general processing speed is the major predictor of positive manifold, the sample was divided into 'slower-than-average' and 'faster-than-average' groups, according to each estimate of inspection time, and the latter correlated with IQ. Considering the first estimate of IT (LIMITS); its correlations with both measures of general intelligence (RPM IQ and global PMA IQ) were appreciably stronger in the slower sample, in line with Anderson's hypothesis. The same result was found using the third estimate (MCS2) and the composite measure. When the sample was divided according to the second IT estimate (MCS1) the correlation of this IT variable with RPM IQ was again stronger in the slower sample, although its relationship with PMA IQ was in the opposite direction, being stronger in the faster group. Although none of the differences were significant at the five percent level (using Fisher's z), all but one were in the predicted direction.

In summary, a general tendency was observed for correlations between IT and IQ to be stronger in samples with slower-than-average IT, compared to those with faster-than-average IT. Although these results do not have the statistical significance necessary to confidently accept the hypothesis, they do provide moderate evidence to support Anderson's theory that mental speed is the determinant of both g-level and g-strength.

<u>Study 4: Using shape rotation/discrimination time as the putative measure of the efficiency of Anderson's spatial-analogical processor (SP2).</u>

Results for the Shape Discrimination Task (Shape RT) replicated the characteristic pattern of increasing response times with an increasing angle of disparity between the two stimulus figures, confirming the validity of the measure. Most IQ variables correlated negatively with average shape discrimination time. The magnitude of RT-IQ correlations was significantly related to the IQ variable's loading on the general factor although (not surprisingly given their similar content) the IQ test showing the strongest correlation with Shape RT was Spatial Relations.

Comparing age groups:

As predicted, twelve year olds had significantly faster average Shape response times than nine year olds. Contrary to the age-differentiation hypothesis of Spearman, Burt and Garrett, however, mean Shape RT correlated *more strongly* with IQ in the *older* of the two samples in eight out of ten cases; the greatest age difference being found for the correlation of mean Shape RT with Raven's Matrices IQ.

Comparing groups differing in ability level:

In 'below-average' and 'above-average' subgroups, divided according to Raven's Matrices IQ, the correlations of RPM IQ and Shape RT were very low, even after correction for range restriction and unreliability, making it difficult to compare the two. Sample division by Primary Mental Abilities IQ, however, produced results in the direction predicted by the ability-differentiation hypothesis, with RT-IQ correlations being non-significantly stronger in the 'below-average' sample than in the 'above average' sample. Nonetheless, study 3 cannot be said to have produced firm evidence to support the hypothesis that RT-IQ correlations will be stronger in samples of lower ability.

Correlations between Shape RT and IQ in groups differing in mental speed: When the sample was divided at the mean according to average *Shape response time*, the correlations between RT and both measures of general intelligence (RPM IQ and PMA IQ) were stronger in the group with 'faster-than-average' RT. Similarly, in groups having 'faster-than-average' and 'slower-than-average' *Inspection Time*, the correlation of Shape RT with both estimates of general intelligence was stronger in the faster sample. Both of these results are in the opposite direction to that predicted, although in neither case was the difference statistically significant. These results do not support Anderson's hypothesis that mental speed determines the strength of g.

Study 5: Using letter identification/discrimination time as the putative measure of the efficiency of Anderson's verbal-propositional processor (SP1).

Results for the letter identification task followed the anticipated pattern whereby the time taken to discriminate between letters on the basis of phonetic similarity was markedly greater than the time required to make a simple visual discrimination. With the exception of Digit Span, all IQ variables correlated negatively with letter discrimination time (Letter RT) in all task conditions. As with Inspection Time and Shape discrimination time, the degree to which average Letter RT correlated with IQ was associated with the IQ tests' g-loadings.

Comparing age groups:

In all conditions, the response times of nine year olds were significantly slower than those of twelve year olds, as hypothesised. Correlations of mean Letter RT with IQ, however, showed no tendency to be either consistently lower or higher in any one age group. General intelligence (estimated from RPM IQ), in fact, correlated more strongly with average Letter RT in the *older* of the two samples. Neither of the latter two results indicates a stronger positive manifold effect in younger groups, consistent with Anderson's theory that the differentiation effect is not age-related.

Dividing the sample according to ability level:

When the sample was divided into two sub-groups according to Raven's Matrices IQ, the correlation of Letter RT with RPM IQ was significantly stronger in the sample of *higher* ability. The opposite result was found, however, when the sample was divided according to global Primary Mental Abilities IQ. Whilst the latter result supports the hypothesis that g is weaker in higher ability groups, the former does not.

When the sample was divided into 'slower-than-average' and 'faster-than-average' sub-groups, on the basis of mean Letter RT, correlations between Letter RT and general intelligence were stronger in the faster of the two groups. Dividing the sample into 'fast' and 'slow' groups according to the most reliable estimate of Inspection Time (MCS2) produced the same pattern of results, with the faster group having higher RT-IQ correlations than the slower group. Study 5 therefore provided no evidence to indicate that mental speed is inversely associated with positive manifold.

Relationship between SP1 and SP2 in groups differing in Inspection Time or IQ.

It had been theorised that Shape RT and Letter RT would tap into the respective efficiencies of Anderson's SP1 and SP2. It was therefore hypothesised that Inspection Time (as the putative measure of the efficiency of the BPM) would predict the degree of relationship between the two RT measures, such that the correlation between mean Shape RT and mean Letter RT would be greater in a sub-group having 'slower-than-average' IT than in a sub-group with 'faster-than-average' IT. Associated with this was the prediction of stronger Shape RT-Letter RT correlations in samples with 'lower-than-average', compared to 'higher-than-average' general intelligence. All of the results pertaining to these comparisons were in the predicted direction, with the correlation between the verbal and spatial RT tasks being stronger in the group with slower, compared to faster IT and lower, compared to higher, IQ. These results are compatible with Anderson's theory that the efficiency of a central, 'basic processing mechanism' predicts the degree to which the outputs of specific *spatial-analogical* (SP2) and *verbal propositional* (SP1) processors are correlated. Nonetheless, since the differences between groups were not statistically significant, these results cannot be accepted as conclusive.

Discussion

The studies reported in this thesis have produced mixed findings and do not provide clear evidence of differences in positive manifold either across the age range studied, or between samples varying in overall ability.

Developmental differences in positive manifold

With respect to the *developmental differences* hypothesised by theorists such as Spearman (1925), Burt (1954), and Garrett (1948); there is no evidence to suggest that psychometric g becomes less pervasive as age increases. Given the lack of consistent evidence to the contrary, Anderson's theory that positive manifold is stable during child development cannot be refuted. However, since the statistical procedures used to compensate for range restriction produced inconsistent effects on the data, it is not possible to be confident of any one interpretation.
The finding that twelve year olds had significantly faster *response times* than nine year olds, is compatible both with Anderson's theory that RT will reflect developmental improvements in knowledge *and* with the theory that RT reflects the speed of central processing resources which increases with age (e.g. Kail, 1991, 1992). With regard to *Inspection Time*, however, the results were inconsistent; some supporting previous findings of improved (i.e. faster) performance with increasing age and others being compatible with Anderson's theory that mental speed (said to be revealed much more clearly by IT than by RT) does not change during childhood. Although results using the most reliable estimate seemed to support Anderson's hypothesis, the fact that the other estimates and the composite measure did not, renders it difficult to draw firm conclusions about the change in mental speed with age. Some light may be shed on this by comparing the impact of the age differences in mental speed on

positive manifold. Nine and twelve year olds did not differ significantly with respect to the size of correlations between Inspection Time and general intelligence, for any IT estimates. Furthermore, the age differences in IT-IQ correlation sizes were smallest for the IT estimate judged to be the most reliable of the three. This implies that Anderson's theory of no age differences in speed and, hence in g-strength, is correct.

Ability group differences in positive manifold

The hypothesis of decreased positive manifold with increasing IQ, postulated by Spearman, Anderson, and others (e.g. Wewetzer,1957) has not been supported. As discussed in chapter 3, however, the true differences between the comparison groups for study 2 were smaller than the selection process implied and the sample sizes were modest. Had the comparison groups been drawn from a larger set of normative data, it may have been possible to select sub-groups who were widely separated on all variables, simultaneously, as in the study by Deary et al. (1996). A comparison of inter-test correlations in mentally retarded and high IQ subjects (e.g. Detterman and Daniel's MR/College study, 1989) would not necessarily be appropriate for studying differences in positive manifold across the spectrum of ability, however, given the possible organic causes of mental retardation.

As expected, Inspection Time and the two Response Time measures correlated negatively with the majority of psychometric test variables, suggesting that, to a greater or lesser extent, they indicate cognitive processing speed. Despite this, there was no overall tendency for either RT or IT to be more strongly correlated with IQ in samples of lower, compared to higher ability. Response times actually tended to correlate more strongly with IQ in the brighter samples; a result which may reflect the shared influence of ability on error in the RT and IQ tasks. None of these results confirm Anderson's and Spearman's prediction that faster mental speed (or 'energy') will be associated with lesser positive manifold. Some evidence in favour of the hypothesis could, however, be gleaned from the nonsignificant tendency for correlations between IQ and IT and between the verbal and spatial RT tasks to be stronger in samples having slower-than-average Inspection Time.

Limitations and suggestions for future research

Methods for estimating inspection time

As discussed in chapter 6, differences between psychophysical methods and difficulties with the estimation of appropriate boundaries for the initiation of psychophysical algorithms, may have introduced error into the measurement of Inspection Times. Further measurement error may, conceivably, have been introduced by the use of the 'space invaders' array, which may have distracted some subjects. It would be useful to repeat the study using a method which is even less prone to extraneous influences.

Cognitive tasks related to SP1 and SP2

In an effort to understand the nature of and relationship between Anderson's SP1 and SP2, it would be beneficial to include a larger number of cognitive tasks related to spatial and verbal ability. This would allow for a clearer comparison of correlations within, versus across, domains at each level of ability and/or general mental speed (IT).

Selection of tests and subjects

There is ample evidence to indicate that range restriction effects may spuriously produce results in line with the differentiation hypothesis, as suspected by Vernon (1965) and others. Even if care is taken to ensure that comparative sub-groups are equal in range on the variable used to select them, it cannot be guaranteed that they will be similarly equal in range on the other variables in the battery from which estimates of g

-strength are derived. A problem is most likely to occur in high-ability samples, since subjects' scores may be close to tests' upper limits, curtailing the effective number of items. Fogarty and Stankov (1996) follow this line of reasoning, attributing the decreasing positive manifold effect to tests' reduced discriminatory power at the highest levels of ability. Variations between tests, with respect to their upper limits, would be expected to exacerbate this spurious differentiation of abilities at higher levels. Although, to the greatest extent, this source of bias has been statistically removed from the data in the studies reported here, it would be useful to repeat the comparisons using a battery of tests with no upper limits.

Heritability of IQ

A potentially fruitful avenue of further research is suggested by recent findings of differential IQ heritability in samples varying in intelligence (Detterman et al, 1990). Anderson (1992) has argued that *"...the person with fast processing is capable of extracting knowledge from the full range of environmental circumstances in which he or she might find him or herself"* (p.130). In other words, brighter individuals will be able to take advantage of a wider range of potential learning experiences than less able individuals. This would imply not only that there should be more variability amongst test scores in high IQ groups (producing a weaker g-factor), but also that the ratio of environmental to genetic influences on those scores should be higher. Findings by Detterman et al (1990) of stronger IQ heritability in low, than in high IQ groups, offer some support for this theory. Other researchers disagree (e.g. Bailey and Revelle, 1991; Cherny et. al., 1992; Thomson et. al., 1993). Nevertheless, an assessment of the relative heritability of IQ in groups varying in ability level may be a useful method of validating the comparisons of such groups with respect to positive manifold.

Conclusions

This author's experiences in attempting to test the differentiation hypotheses lead her to the same conclusion reached by Horn (1970), namely that *"it is difficult to ask questions about differentiation in ways that make the questions answerable by empirical research."* (p.425) Similarly, Rienert (1970) remarked upon the difficulty of evaluating the widely inconsistent results of studies up to 1969, *"given the complexity of the research object".* Perhaps the main contribution of this research has been to analyse the complexity to which Reinert referred, particularly as regards disentangling the effects of statistical bias from true differences in positive manifold. Research for study 1, for example, uncovered a general dearth of knowledge, amongst researchers in the field, of procedures which can be applied to the problem of range restriction in samples which have relatively narrow distributions, but have not been explicitly selected from a broader sample with known characteristics. Study 2 drew attention to the fact that selecting groups on the basis of one variable, does not guarantee that they will be as widely separated or homogeneous, with respect to other abilities. Whilst this fact is established in fundamental psychometric theory, it has often been overlooked in previous work and may account for the inconsistent findings of different investigators who may have used more or less pervasive variables to select their comparative ability ranges.

No consistent relationship has been found between age, ability, or 'mental speed' and the magnitude of positive manifold. It must therefore be concluded that the 'reducing positive manifold effect', where it has been found, results from something other than differences in fundamental processes associated with individual or developmental differences in mental efficiency. A more plausible explanation is the one given by Carrol (1993) and others before him (e.g. Cattell, 1971), in regard to age differences: *"To the extent that differentiation actually occurs, it centres on abilities having to do with different areas of learning and skill formation"* (p.687). This argument can be extrapolated to ability group differences, since it is reasonable to assume that higher ability is associated with greater opportunities for learning and specialisation of interests. Of course, such a claim cannot be refuted with the evidence available from the investigations reported here. Nevertheless, the evidence derived from these investigations suggests that the phenomenon, if it occurs, is not a consequence of internal facets of cognitive architecture, but is more likely to stem from environmental (or interactive) factors.

Closing remarks

This thesis has been equally concerned with uncovering the history of the differentiation hypotheses, as with testing the various predictions arising from Anderson's theory. Although the studies reported have not yielded any clear evidence to indicate that positive manifold decreases with increasing age or IQ level, they do reveal a number of extraneous factors which can influence it. The author hopes to have played some part in acquainting readers with a 'lost' literature and with re-establishing the notion that Spearman's g may not be uniformly influential across the ability range. Further research may help to clarify some of the factors which can affect the strength of g, and this may be useful for advancing the science of psychometric testing. It is difficult, however, to envisage a clear resolution of this problem given the multiplicity of factors which may interact in producing sample differences in g and the failure of so many previous studies to produce consistent results.

REFERENCES

Ackerman, P.L. (1986) Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence*, 10, 101-139

Ackerman, P.L. (1987) Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102* (1), 3-27.

Anastasi, A. (1958,1964) *Differential Psychology.* N.Y. MacMillan

Anastasi, A. (1970) On the formation of psychological traits. *American Psychologist, 25,* 899-910

Anderson, M. (1986) Understanding the cognitive deficit in mental retardation. *Journal of Child Psychology and Psychiatry, 27,* 3, 297-306

Anderson, M. (1987) The myth of multiple intelligences. Paper presented to British Psychological Society, Cognitive Section: Annual Conference, September 1987.

Anderson, M. (1989a) New ideas in intelligence. *Psychological Bulletin,3*, 92-94

Anderson, M. (1989b) The effect of attention on developmental differences in inspection time.*Personality and Individual Differences,* 5, 559-563

Anderson, M. (1989c) The cognitive basis of intelligence: Is Triarchic theory a hostage to fortune? (unpublished manuscript)

Anderson, M. (1992) *Intelligence and Development: A Cognitive Theory.* Oxford: Blackwell

Anderson, M. (1988) Inspection time, information processing and the development of intelligence. *British Journal of Developmental Psychology, 6*, 43-57

Anderson, H.E. & Leton, D.A.(1964) Factor analysis of the California Test of Mental Maturity. *Educational and Psychological Measurement, 24,* 513-523.

Asch, S. (1936) A study of change in mental organization. *Archives of Psychology*, 195, 17-24.

Atkin, R., Bray, R., Davidson, M., Herzberger, S., Humphreys, L. & Selzer, U. (1977). Ability factor differentiation, grades 5 through 11. *Applied Psychological Measurement, 1*, 1, 65-77.

Baddeley, A.D. & Hitch, G. (1976) Working Memory. IN: G.Bower (Ed.) *Advances in learning and motivation.* N.Y.: Academic Press.

Bailey, M.J. & Revelle, W. (1991) Increased heritability for lower IQ levels? *Behavior Genetics, 21,* 397-404

Balinsky, B. (1941) An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs, 23,* 191-234.

Barron, F.(1963) *Creativity and psychological health.* Princeton, N.J.: Van Nostrand.

Beech, J.R. & Harding, L. (1990) *Testing People: A practical guide to psychometrics.* Windsor: NFER-NELSON

Binet, A. (1905) Analyse de C.E. Spearman, the proof and measurement of association between two things and general intelligence objectively determined and measured. *L'Annee Psychologique, 11,* 623-624

Bickley, P.G., Keith, T.Z. & Wolfle, L.M. (1995) The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the lifespan. *Intelligence, 20,* 3, 309-328.

Binet, A (1908) Le development de l'inteligence chez les enfants. *L'Annee Psychologique, 14*, 1-94

Binet, A. (1911) Nouvelles recherches sur la mesure du niveau intellectual chez les enfants d'ecole. *L'Annee Psychologique, 17*, 145-201

Binet, A., & Henri,V. (1895) La psychologie individuelle. *L'Annee Psychologique, 2*, 411-465

Binet, A.,& Simon, T.(1905) Methodes nouvelles puor le diagnostic du niveau intellectuel des anormaux. *Annee psychol., 11*, 191-244

Bouyer, S. & Kneip, N. (1981) Structure mentale et Q.I. Contribution a l'hypothese de divergence de Wewetzer. *Revue de Psychologie Appliquee, 31* (3), 169-178.

Brand, C. (1981) General intelligence and inspection time: Their relationship and development. IN: M.P.Friedman, J.P.Das, & N.O'Connor (Eds.) *Intelligence and Learning.* N.Y.: Plenum

Brand, C.R. (1984) Intelligence and inspection time: An ontogenetic relationship? IN: C.J.Turner & H.B.Miles (Eds.) *The biology of human intelligence: Proceedings of the twentieth annual symposium of the eugenics society. London, 1983.* U.K.: Nafferton books

Brand, C.R. & Deary, I.J. (1982) Intelligence and "inspection time". IN: H.J.Eysenck (Ed.) *A model for intelligence.* Berlin & N.Y.:Springer-Verlag

Brody, N. (1992) *Intelligence.* London: Academic Press

Burt, C. (1919) The bearing of the factor theory on the organization of schools and classes. Report of the L.L.C. Psychologist reprinted in King (1921) *Mental and Scholastic Tests*

Burt, C. (1949) The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology, 19,* 100-111, 176-199.

Burt, C. (1954) The differentiation of intellectual ability. *British Journal of Educational Psychology, 24,* 79-90

Butcher, H.E (1968) *Human Intelligence - Its Nature and Assessment.* London: Methuen

Carroll, J.B. (1941) A factor analysis of verbal abilities. *Psychometrika, 6,* 279-307.

Carroll, J.B. (1993) *Human Cognitive Abilities: A survey of factor-analtyic studies.* Cambridge University Press.

Cattell, J. McK. (1850) Mental tests and measurements. *Mind,* 15, 373 -380

Cattell, R.B. (1941) Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R.B. (1966) The Scree test for the number of factors. *Multivariate Behaviour Research, 1,* 140-161.

Cattell, R.B. (1967) The theory of fluid and crystallised intelligence. *British Journal of Educational Psychology, 37*, 209-224.

Cattell, R.B. (1971) *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin

Cattell, R.B. (1966) Higher order factor structures and reticular vs hierarchical formulae for their interpretation. IN: C.Banks & P.L. Broadhurst (Eds.), *STEPHANOS: Studies in psychology.* N.Y.: Barnes & Noble.

Cattell, R.B. (1978) *The Scientific Use of Factor Analysis in Behavioral and Life Sciences.* New York: Plenum.

Chen, T.L. & Chow,H.A. (1948) A factor study of a test battery at different educational levels. *Jn. Genet.Psychol.,73,* 187-199

Cherny, S.S., Cardon, L.R., Fulkner, D.W., & DeFries, J.C. (1992) Differential heritability across levels of cognitive ability. *Behavior Genetics, 22,* 153-1692

Chrysostum, F. (1938) L'organizaton des traits mentaux avec l'age. *Ann. Assoc. Canadienne-Francais pour l'Avancement des Sciences*, iv 124-6

Chuprikova, N.I. (1990) Principle of differentiation of cognitive structures in mental development, education and intelligence. *Voprosy Psikhologii, 5,* 31-39.

Clark, M.P. (1944) Changes in mental organization with age. *Archives of Psychology,291,* 30

Corballis, M.C. (1989) Laterality and human evolution. *Psychological Review 96*, 3, 492-505.

Cooper, L.A. and Shepard, R.N. (1973) Chronometric studies of the rotation of mental images. In W.G. Chase (Ed.) *Visual information processing.* New York: Academic Press

Cronbach, L.J. (1973) *Essentials of Psychological Testing (4th Ed.)* New York: Harper & Row.

Cropely, A.J. (1964) Differentiation of abilities, socioeconomic status, and the WISC. *Journal of Consulting Psychology, 28*, 512-517

Cropley, A.J. (1966) Creativity and intelligence. *British Journal of Educational Psychology, 36,* 259-266.

Curtis, H.A. (1950) A study of the effects of age and test difficulty on factor patterns. *Genetic Psychology Monographs.*

Deary, I.J. & Stough, C. (1996) Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51* (6), 599-608.

Deary, I.J.; Egan, V; Gibson, G.J.; Austin, E.J.; Brand, C.R.; Kellaghan, T. (1996) Intelligence and the Differentiation Hypothesis. *Intelligence, 23*

Deary, I.J. & Pagliari, C. (1991) The strength of g at different fevels of ability: Have Detterman and Daniel rediscovered Spearman's "law of diminishing returns"? *Intelligence, 15* (2), 247-250.

Detterman, D.K. (1987) Theoretical notions of intelligence and mental retardation. *American journal of mental deficiency, 92* (1) 2-11.

Detterman, D.K. (1991) Reply to Deary and Pagliari: Is g intelligence or stupidity? *Intelligence, 15* (2) 251-255.

Detterman, D.K. & Daniel, M.H. (1989) Correlations of Mental Tests with Each Other and with Cognitive Variables are Highest for Low IQ Groups. *Intelligence, 13,* 349-359

Detterman, D.K., Mayer, J.D., Caruso, D.R; Legree, P.J., et al. (1992) Assessment of basic cognitive abilities in relation to cognitive deficits. *American Journal on Mental Retardation, 97* (3), 251-286

Detterman, D.K., Thompson, L.A. & Plomin, R. (1990) Differences in heritability across groups differing in ability. *Behavior Genetics, 20,* 369-384.

Doppelt, J.E. (1949) The organization of mental ability in the age-range thirteen to seventeen. *American Psychologist, 4,* 242-253

Dunlap, W.P., Chen, R. & Greer, T. (1994) Skew reduces test-retest reliability. *Journal of Applied Psychology, 79,* 2, 310-313.

Eberle, G., Holtz, K-L., Kowalewski, A., Staiger, M. (1979) A study to clarify the factorial structure of the Psycholinguistic Development Test in so-called learning-disabled special school students. *Zeitschrift fur Entwicklungspsychologie und Pedagogische Psychologie, 10* (2), 134-143.

Ekstrom, R.B., French, J.W., & Harman, H.H. (1976) *Manual for kit of factor referenced tests.* Princeton, N.J. Educational testing service.

Elliott, C.D. (1983) *British Ability Scales. Manual 1: Introductory handbook. Manual 2: Technical handbook.* Windsor, England: NFER/Nelson

Eysenck, H.J. (1939) Review of Thurstone's Primary Mental Abilities, 1938. *British Journal of Educational Psychology, 9,* 270-275.

Eysenck, H.J. (1973) *The Measurement of Intelligence.* Medical & Technical Publishing Co. Ltd.

Eysenck, H.J. (1988) The concept of 'intelligence': useful or useless? *Intelligence, 12,*1-16.

Fagan, J.F. & McGrath, S.K. (1981) Infant recognition memory and later intelligence. *Intelligence, 5,* 121-130.

Flynn, J.R. (1987) Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171-191

Fillela, J.F.(1957) Educational and sex differences in the organization of abilities in technical and academic students in Columbia, South America. Unpublished doctoral dissertation, Fordham University, 1957.

Fodor, J. (1983) *The modularity of mind.* Cambridge Massachusetts: MIT Press.

Fogarty, G.J., & Stankov, L. (1995) Challenging the "Law of Diminishing Returns". *Intelligence, 21,* 157-174

Frearson, W.M., Barrett, P; & Eysenck, H.J. (1988) Intelligence, reaction time and the effects of smoking. *Personality and individual differences, 9,* 497-519

Galton, F. (1869) *Hereditary Genius: an inquiry into its laws and consequences.* N.Y. MacMillan

Galton, F. (1887) Notes on prehension in idiots. *Mind 12,* 79-82

Gardner, H.E. (1983) *Frames of Mind: The Theory of Multiple Intelligences.* London: Methuen

Garret, H.E. (1930) A study of the CAVD intelligence examination. *J.Educ.Res.,21*, 103-108

Garrett, H.E. (1938) Differentiable mental traits. *Psychological Record, 2,* , 259-298.

Garret, H.E. (1946) A Developmental Theory of Intelligence. *American Psychologist, 1*, 372-378

Garret, H.E., Bryan, A.I. & Perl, R. (1935) The age factor in mental organization. *Archives of Psychology, 176*, 31

Getzels, J.W & Jackson, P.W. (1962) *Creativity and Intelligence.* N.Y.: Wiley

Goodman, J.F. & Cameron, J. (1978) The meaning of IQ constancy. *Journal of Genetic Psychology 132,* 109-119

Guilford, J.P. (1950) Creativity. *American Psychologist*, *5*, 444-454

Guilford, J.P. (1967) *The Nature of Human Intelligence.* N.Y. McGraw-Hill

Guilford, J.P. (1982) Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review*, *89*, 48-59

Guilford, J.P. & Hoepfner,R. (1971) *The analysis of intelligence.* New York: McGraw-Hill.

Gulliksen, H. (1950) *Theory of Mental Tests.* New York: Wiley.

Gustaffson, J-E. (1984) A Unifying Model for the Structure of Intellectual Abilities.*Intelligence*, *8*, 179-203

Gustaffsson (1988) Hierarchical models of individual differences. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, NJ: Erlbaum

Hargreaves, D.J. & Bolton, N. (1972) Selecting Creativity Tests For Use in Research. *British Journal of Psychology, 63*, 3, 451-462

Harman, H. (1967) *Modern factor analysis.* Chicago: University of Chicago Press.

Hasan, P. & Butcher, H.J.(1966) Creativity and intelligence: a partial replication with Scottish children of Getzels and Jackson's study. *British Journal of Psychology, 57*, 129-135.

Hendrickson, A.E. (1982) The biological basis of intelligence. Part I: Theory. IN: H.J.Eysenck (Ed.) *A model for intelligence.* Berlin & N.Y.: Springer-Verlag

Hendrickson, D.E. (1982) The biological basis of intelligence. Part II: measurement. IN: H.J.Eysenck (Ed.) *A model for intelligence.* Springer-Verlag

Hendrickson, D.E. & Hendrickson, A.E. (1980) The biological basis of individual differences in intelligence. *Personality and Individual Differences,1*, 3-33

Hermelin, B. & O'Connor, N. (1983) The Idiot-Savant : Flawed Genius or Clever Hans. *Psychological Medicine,13*, 479-481

Hermelin, B. & O'Connor, N. (1986) Idiot-savant calendrical calculators: rules and regularities. *Psychological Medicine, 16,* 885-893

Horn, J.L. (1968) Organization of abilities and the development of intelligence. *Psychological Review, 79,* 242-259

Horn, J.L. (1970) Personality and ability theory. In R.B.Cattell (Ed.), *Handbook of modern personality theory.* New York: Aldine

Horn, J. & Cattell, R.B. (1966) Refinement and test of the theory of fluid and crystallised intelligence. *Journal of Educational Psychology, 57,* 253-270.

Hosie, B.M. (1979) Mental speed and intelligence: Their relationship and development in 4-year-old children. Unpublished undergraduate thesis, University of Edinburgh.

Hsu, E.H. (1948) Factor analysis, differential bio-process, and mental organization. *Journal of Genetic Psychology,38*, 147-157

Hudson, L. (1965) Intelligence: convergent and divergent. In: *Penguin Science Survey*,1965.

Hudson, L. (1966) *Contrary Imaginations.* London: Methuen.

Hulme, C. & Turnbull, J. (1983) Intelligence and inspection time in normal and mentally retarded subjets. *British Journal of Psychology,74*,365-370

Humphreys, L.G. (1982) The hierarchical factor model and general intelligence. IN: N. Hirschberg & L.G. Humphreys (Eds.) *Multivariate applications in the social sciences.* Hillsdale N.J.: Erlbaum.

Hunt, E. (1978) Mechanics of verbal ability. *Psychological Review, 85*, 109-130.

Hunt, E.B. (1980) Intelligence as an information processing concept. *British Journal of Psychology, 7*, 449-474.

Hunt, E., Frost,N., & Lunneborg,C. (1973) Individual differences in cognition: A new approach to intelligence. IN: G.H.Bower (Ed.) *The Psychology of Learning and Motivation.*(vol.7) N.Y.: Academic Press

Hunt, E.B., Lunneborg, C. & Lewis, J. (1975) What does it mean to be high verbal? *Cognitive Psychology, 7,* 194-227.

Hurst, J.G. (1960) A factor analysis of the Merrill-Palmer with reference to theory and test constuction. *Educational and Psychological Measurement, 20*, 519-532

Irwin, R.J. (1984) Inspection time and its relation to intelligence. *Intelligence,8*,47-65.

Jensen, A.R. (1982) Reaction time and psychometric g. IN: H.J. Eysenck(Ed.) *A Model For Intelligence.* Berlin & New York: Springer-Verlag.

Jensen, A.R. (1987) Process differences and individual differences in some cognitive tasks. *Intelligence, 11*, 107-136.

Jones, L.V. (1949) A factor analysis of the Stanford-Binet at four age levels. *Psychometrika, 14,* 299-331

Joreskog, K.G. (1963) *Statistical estimation in factor analysis.* Uppsala, Sweden: Almqvist & Wiksells.

Kail, R. (1991) Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490-501.

Kail, R. (1992) Evidence for global developmental change is intact. *Journal of Experimental Child Psychology, 54* (3), 308-314.

Kaiser, H.F. (1968) A measure of the average intercorrelation. *Educational and Psychological Measurement, 28*, 245-247.

Keating, D.P., & Bobbit, B.L. (1978) Individual and developmental differences in cognitive-processing components of mental ability. *Child Development*, *49*, 155-167

Kelly, T.L. (1928) *Crossroads in the mind of man: A study of differentiable mental abilities.* Stanford California: Stanford

Kline, P. (1993) *The Handbook of Test Construction.* London: Methuen.

Kline, P. (1991) *Intelligence: The Psychometric View.* Lodon: Routledge

Kline, P. (1993) *The Handbook of Psychological Testing.* London: Routledge.

Kosslyn, S.M. (1980) *Image and Mind.* Harvard Univ. Press.

Kosslyn, S.M. (1981) The medium and the message in mental imagery: a theory. *Psychological Review 88*, 46-66

Kosslyn, S.M. (1983) Ghosts in the Mind's Machine: Creating and Using Images in the Brain. New York: W.W.Norton.

Kranzler, J.H. & Jensen, A.R. (1989) Inspection time and intelligence: A meta-analysis. *Intelligence, 13,* 329-48.

Kranzler, J.H. & Jensen, A.R. (1991) The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence 15,* 297-422.

Lally, M. & Nettlebeck,T. (1977) Intelligence, reaction ime, and inspection time. *American Journal of Mental Deficiency,82*, 273-281.

Lawley, D.N. (1943) A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62,* 28-30.

Legree, P.J., Pifer, M.E. & Grafton, F.C. (1996) Correlations among cognitive abilities are lower For higher ability groups. *Intelligence, 23*, 45-57

Levine, G., Preddy, D., & Thornike, R.L. (1987) Speed of information processing and level of cognitive ability. *Personality and Individual Differences, 8*, 5, 599-607

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. *Journal of the Acoustics Society of America, 49*, 467-477.

Lienert, G.A. (1960) Die faktorenstruktur der intelligenz als funktion des intelligenzniveaus. In H. Thomae (Ed.), *Bericht uber den XXII. Kongress der Deutschen Gesellschaft fur Psychologie in Heidelberg, 1959.* Gottinger: Hogrefe. (1959, 1960)

Lienert, G.A. & Crott, H.W. (1964) Studies on the factor structure of intelligence in children, adolescents and adults. *Vita Humana, 7,* 147-163.

Lienert, G.A. & Faber, C. (1963) Uber die factorenstruktur des HAWIK auf verschiedenen alterund intelligenzniveaus. *Diagnostica, 9,* 3-11.

Longstreth, L.E. (1984) Jensen's reaction time investigations of intelligence: A critique. *Intelligence, 8,* 139-160.

Lynn, R. (1987) The intelligence of the mongoloids: a psychometric, evolutionary and neurological theory. *Personality and Individual Differences, 8*, 813-844.

Lynn, R. (1992) Does Spearman's g decline at high levels? Some evidence from Scotland. *Journal of Genetic Psychology, 153, 229-230*

Lynn, R. & Cooper, C. (1993) A secular decline in Spearman's g in France. *Learning and Individual Differences, 5,* 43-48

Lynn, R. & Cooper, C. (1994) A secular decline in Spearman's g in Japan. *Current Psychology: Developmental, Learning, Personality, Social, 13*, 3-9

Lynn, R., Pagliari, C. & Chan, J. (1988) Intelligence in Hong Kong measured for Spearman's g and the visuospatial and verbal primaries. *Intelligence, 12*, 423-433

MacKinnon, D.W. (1962) The nature and nurture of creative talent. *Amer. Psychologist, 17*, 484-495

Mackintosh, N.J. (1981) A new measure of intelligence? *Nature, 289*, 529-530.

McCall, R.B. & Carriger, M.S. (1993) A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development, 64* (1), 57-79.

McCartin, Sister Rose Amata, & Meyers,C.E. (1966) An exploration of six semantic factors at first grade. *Multivariate Behavioural Research, 1*, 74-94

McClelland, D.C. (1958) Issues in the identification of talent IN *Talent and society* (D.C McClelland ed.)N.Y. : Appleton-Century-Crofts.

McGrew, K.S., Werder, J.K. & Woodcock, R.W. (1991) *WJ-R technical manual.* Allen, Texas: D.L.M.

Meer, B., & Stein, M.I. (1955) Measures of inteligence and creativity. *Journal of Psychology,39*, 117-126

Micceri, T. (1989) The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Mitchell, J.V. Jr. (1956) A comparison of the factorial structure of cognitive functions for a high and low status group. *Journal of Educational Psychology, 47,* 397-414.

Moore, R. (1967) The relation of intelligence to creativity. *Journal of Research in Education, 14,* 143-253

Murphy, K.R. & Davidshofer, C.O. (1991) *Psychological Testing: Principles and Applications (2nd ed.).* Englewood Cliffs, New Jersey: Prentice Hall.

Nettelbeck, T. (1982) Inspection time: an index for intelligence? *Quarterly Journal of Experimental Psychology, 34A,* 299-312.

Nettelbeck, T. (1987) Inspection Time and Intelligence. IN : P.A. Vernon (ed.) *Speed of Information-Processing and Intelligence.* N.J. Ablex.

Nettelbeck,T. & Kirby, N.H. (1983a) Measures of timed performance and intelligence. *Intelligence, 7*, 39-52

Nettelbeck, T. & Kirby, N.H. (1983b) Retarded-nonretarded differences in speed of processing. *Australian Journal of Psychology, 35*, 445-453

Nettelbeck, T. & Lally,M. (1976) Inspection time and measured intelligence. *British Journal of Psychology, 67,* 17-22

Nettelbeck, T. & Wilson, C. (1994) Childhood changes in speed of information processing and mental age: A brief report. *British Journal of Developmental Psychology, 12* (3), 277-280.

Nettelbeck, T. & Wilson, C. (1985) A cross-sequential analysis of developmental differences in speed of visual information processing. *Journal of Experimental Child Psychology, 40,*1-22

Nettelbeck, T. & Vita, P. (1992) Inspection time in two childhood age cohorts: A constant or developmental function? *British Journal of Developmental Psychology, 10* (2), 189-197.

Norman, D.A. & Bobrow, D.B. (1975) On data-limited and resource-limited processes. *Cognitive Psychology,7,* 44-64.

O'Connor, N. & Hermelin, B. (1983) The role of general ability and specific talents in information processing. *British Journal of Developmental Psychology, 1,* 389-403.

O'Neill, W.M. (1962) The stability of the main patterns of abilities with changing age. *Australian Journal of Psychology,14,* 1-8

Oerter,R., Mandl, H., & Zimmermann, A. (1974) New findings concerning the differentiation hypothesis of intelligence. *Zeitschrift fur Entwicklungspsychologie und Pedagogische Psychologie. 6 (3),* 151-167

Paraskevopolous, J.N. & Kirk, S.A.(1969) *The development and psychometric characteristics of the revised Illinois Test of Psycholinguistic Abilities.* Urbana,IL: University of Illinois Press.

Pearson, K. (1903) Mathematical contributions to th theory of evolution: II. On the influence of natural selection on the variability and correlation of organs. *Royal Society Philosophical Transactions, 200,* 1-66.

Peel, E.A. & Graham, D. (1951-52) The differentiation of ability in primary school children. *Research review, 2* ,40-48. 3, 31-34

Piaget, J. (1950) *The psychology of intelligence.* N.Y.: Harcourt, Brace, & World.

Piaget, J. (1953) *The Origin of Intelligence in the Child.* Routledge.

Posner, M.I. & Mitchell, R.F. (1967) Chronometric analysis of classification, *Psychological Review, 74,* 392-409

Rabbit, P.M.A. (1985) Oh g Dr.Jensen! or, g-ing up cognitive psychology? Open peer commentary. *Behavioral and Brain Sciences,8*, 238-239.

Ree, M.J, Caretta, T.R; Earles, J.A., & Albert, W. (1994) Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology, 79* ,(2), 298 -301.

Reiben, L. & Mengal, P. (1977) Intelligence globale, creativite et operativite chez l'enfant: Analyse factorielle et analyse discriminante. [Global intelligence, creativity, and operativity in the child: Factorial and discriminant analysis.] *Psychologie - Schweizerische Zeitschrift fur Psychologie und ihre Anwendungen, 36*, 100-108.

Reichard, S. (1944) Mental organization and age level. *Archives of Psychology, 295*

Reinert, G. (1964) Zur problematik der faktoriellen differenzierungshypothese der intelligenz. In H. Heckhausen (Ed.), *Bericht uber den 14. Kongress der Deutschen Gesellschaft fur Psychologie in Wien.* Gottingen: Hogrefe.

Reinert, G. (1970) Comparative factor analytic studies of intelligence throughout the human lifespan span. IN: L.R. Goulet & P.B. Baltes (Eds.), *Life-span developmental psychology: Research and theory.* (pp.467-484). New York: Academic Press.

Reinert, G., Baltes, P., & Schmidt, L.R. (1965) Faktoranalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: Die Leistungsdifferenzierungshypothese. [Factor-analytic investigations of the hypothesis of intelligence differentiation: The hypothesis of differentiation by performance level.] *Psychologische Forschung, 28,* 246-300.

Richards, T.W. & Nelson, V.L. (1939) Abilities of infants during the first eighteen months. *Journal of Genetic Psychology, 55,* 299-318

Rose, D.H., Slater, A.S., & Perry, H. (1986) Prediction of childhood intelligence from habituation in early infancy. *Intelligence, 10*, 251-263

Rust, J. & Golombok, S. (1989) *Modern psychometrics: The science of psychological assessment.* London: Routledge.

Schneider, W. & Schiffrin, R.M. (1977) Automatic and controlled processing and attention. *Psychological Review, 84*, 1-66.

Sharma, S. & Triptish, A. (1990) Antecedents of psychological differentiation. *Journal of Personality and Clinical Studies, 6*, 2, 185-192.

Schaie, K.W. & Herzog, C. (1986) Toward a comprehensive model of adult intellectual development: Contributions of the Seattle Longitudinal Study. In. R.J.Sternberg (Ed.), *Advances in the psychology of human intelligence,* Vol.3. (pp.79-118) Hillsdale. NJ: Erlbaum.

Shepard, R.N. & Feng, C. (1972) A Chronometric Study of Mental Paper Folding. *Cog. Psych. 3*, 228-243

Shepard, R.N. & Metzler, J. (1971) Mental rotation of three -dimensional objects. *Science 171,* 701-703

Shiller, B. (1934) Verbal, numerical and spatial abilities in young children. *Archiv. Psychol. 161,* 69

Schultz, N.R., Kaye, D.B., & Hoyer, W.J. (1980) Intelligence and spontaneous flexibility in adulthood and old age. *Intelligence, 4,* 219-231.

Smith, O.W. & Smith, P.C. (1966) Developmental studies of spatial jugements by children and adults. *Perceptual and Motor skills, 22,* 3-73

Smith, G. & Stanley, G. (1983) Clocking g: Relating intelligence and measures of timed performance. *Intelligence,7*, 353-68.

Snow, J.H. & Strope, E.E. (1990) Development of mental rotation matching abilities with children. *Developmental Neuropsychology, 6* (3), 207-214.

Spearman, C. (1904) "General intelligence" objectively determined and measured. *Amer. Journ. Psychol.,15*, 201-293

Spearman.C. (1923) *The nature of "intelligence" and the principles of cognition.* London: Macmillan.

Spearman, C.E. (1925) *Some Issues in the Theory of "G" (Including the Law of Diminishing Returns)* From the collected papers of the psychological laboratory, University College London 1925.

Spearman, C.E. (1927) *The Abilities of Man.* N.Y. MacMillan.

Spearman, C. & Holzinger, K.J. (1924) The sampling error in the theory of two factors. *British Journal of Psychology, 15*, 17-19

Spencer, H. (1870) *The Principles of Psychology* (2nd. ed). N.Y. MacMillan

Spitz, H.H. (1982) Intellectual Extremes, Mental Age, and the Nature of Human Intelligence. *Merril-Palmer Quarterly, 28*, 2, 167-192

Spitz, H.H. (1986) *The Raising of Intelligence.* London: Earlbaum

Stern, W. (1921) *Uber Psychologie der individuellen Differenzen.* Leipsig : Barth

Sternberg, R.J. (1977) *Intelligence, information-processing and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum

Sternberg, R.J. (1980) Sketch of a componential subtheory of human intelligence. *Behavioral and Brain Sciences, 3*, 573-614

Sternberg, R.J. (Ed.) (1982) *Handbook of Human Intelligence.* Cambridge Univ. press.

Sternberg, R.J. (1983) Components of human intelligence. *Cognition, 15*, 1-48.

Sternberg, R.J. (1984) Toward a triarchic theory of human intelligence. *Behavioral and Brain Sciences 7* (2) 269-315

Sternberg, R.J. (1985) *Human Abilities - An Information Processing Approach.* N.Y. W.H. Freeman & Co.

Sternberg, R.J. (1985) *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R.J. (1988) Explaining away intelligence: A reply to Howe. *British Journal of Psychology, 79*, 527-533.

Sternberg, S. (1965) High speed scanning in human memory. *Science, 153*, 652-654

Sternberg, S. (1970) Memory scanning; Mental processes revealed by reaction time experiments. IN J. Antrobus (Ed.) *Cognition and affect.* Boston: Little,Brown.

Sternberg, S. (1975) Memory scanning: new findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1-32

Stott, L.H. & Ball, R.S. (1963) *Evaluation of infant and preschool mental tests.* Detroit, Michigan : Merrill-Palmer

Stricker, L.J. & Rock, D.A. (1987) Factor structure of the GRE General Test in young and middle adulthood. *Developmental Psychology, 23* (4), 526 -536.

Sullivan, J. (1973) The relationship of creative and convergent thinking to literal and critical reading ability of children in the upper grades. *Journal of Educational Research, 66,* 374-377.

Sumita, K. & Ichitani, T.A. (1958) A factor analyic study on the differentiation of intellectual abilities. *Tohoku Psychologica Folia, 16,* 51-85

Swineford, F. (1949) General, verbal and spatial bifactors after three years. *Journal of Educational Psychology, XL* 353-360

Terman, L.M. (1916) *The measurement of intelligence.* Boston: Houghton-Mifflin

Terman, L.M. & Merrill, M.A. (1960) *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M.* Boston: Houghton Mifflin.

Thompson, L.A., Detterman, D.K., & Plomin, R. (1993) Differences in heritability across groups differing in ability, revisited. *Behavior Genetics, 23,* 331-336

Thomson, G.H. (1916) A hierarchy without a general factor. *British Journal of Psychology, 8*, 271-281

Thorndike, R.L. (1927) *The Measurement of intelligence.* N.Y.: Teachers Coll.

Thorndike, R.L. (1949) *Personnel Selection.* New York: Wiley.

Thorndike, R.L. (1994) g. *Intelligence 19(2)* 145-155

Thurstone, L.L. (1931) Multiple factor analysis. *Psychological Review, 38*, 406-427.

Thurstone, L.L. (1938) Primary Mental Abilities. *Psychometric monographs, 1*

Thurstone, L.L. (1955) Differential Growth of Mental Abilities. *Science, 121*, 627.

Thurstone, L.L. & Thurstone, T.G. (1941) Factor studies of intelligence. *Psychometric monographs, 94*

Thurstone, T.G. (1963) *Examiner's Manual for the Primary Mental Abilities Test for Grades 4-6.* Chicago: Science Research Associates.

Torrance, E.P. (1962) *Guiding creative talent.* Englewood Cliffs, NJ: Prentice-Hall.

Toussaint, N.A. (1974) An analysis of synchrony between concrete -operational tasks in terms of structure and performance demands. *Child Development, 45,* 992-1001.

Tylor-Wood, T. & Carri, L. (1991) Identification of gifted children: The effectiveness of various measures of cognitive ability. *Reoper Review, 14,* 2, 63-64.

Undheim, J.O. (1979) Broad ability factors in 12 to 13 year old children: The theory of fluid and crystallized intelligence and the differentiation hypothesis. *Journal of Educational Psychology 70*, 3, 433-443.

Vernon, P.A. (1983) Speed of information processing and general intelligence. *Intelligence,7,* 53-70.

Vernon, P.A. (1987) *Speed of Information Processing and Intelligence.* N.J.: Ablex

Vernon, P.E. (1965) *The Structure of Human Abilities* (3rd ed.) London: Methuen

Vernon,P.E. & Parry, J.B. (1949) *Personnel selection in the British Forces.* London: University of London Press.

Vickers, D. (1970) Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*, 37-58

Vickers, D. Nettelbeck, T. & Willson, R.J. (1972) Perceptual indices of performance: The measurement of "inspection time" and "noise" in the visual system. *Perception, 1,* 263-295.

Wachs, T.D. & Hubert, N.C. (1981) Changes in the structure of cognitive-intellectual performance during the second year of life. *Infant Behavior and Development 4,* 151-161.

Wallach, M. & Kogan, N. (1965) A new look at the creativity - intelligence distinction *Journal of Personality,33*, 348-369

Wallbrown, F.H., Blaha, J. & Wherry, R.J. (1973) The hierarchical factor structure of the Wechsler Preschool and Primary Scale of Intelligence. *Journal of Consulting and Clinical Psychology, 41* (3), 356-362

Weschler, D. (1958) *The measurement and appraisal of adult intelligence.* (4th. ed.) Baltimore: Williams & Wilkin

Wechsler, D. (1974) *Manual for the Wechsler Intelligence Scale For Children - Revised.* New York: The Psychological Corporation.

Werdelin, I. & Stjernberg, G. (1995) Age differences in factorial structure: A study of the "differentiation hypothesis". *Interdisciplinaria, 12* (2), 79-97.

Wewetzer, K.H. (1958) Zur differenz der leitungsstrukturen bei verschiedened intelligenzgraden. In A. Wellek (Ed.), *Bericht uber den 21. Kongress der Deutschen Gelleschaft fur Psychologie in Bonn, 1957.* Gottingen: Hogrefe.

Williams, H.S. (1948) *Some aspects of the measurement and maturation of mechanical aptitude in boys aged twelve to fourteen.* Unpublished thesis. University of London.

Wilson, C. (1984) *Developmental studies in timed performance.* Doctoral dissertation, University of Adelaide, South Australia.

Wilson, C; Nettelbeck, T, Turnbull, C, Young, R. (1992) IT, IQ and age: A comparison of developmental functions. *British Journal of Developmental Psychology, 10* (2), 179-188.

Wilson, C. & Nettlebeck, T. (1986) Inspection Time and the Mental Age Deviation Hypothesis. *Personality and Individual Differences.,7*, 5, 669-675

Wiseman, S. (1964) *Education and environment.* Manchester: Manchester university press.

Wissler, C. (1901) The Correlation of mental and physical traits. *Psychological Monographs, 3,* (6, Whole No. 16).

Woodcock, R. & Johnson, M.B. (1977-8) *Woodcock-Johnson Psycho-Educational Battery.* Hingham, MA: Teaching Resources

Yamamoto, K. (1964) Threshold of intelligence in academic acheivement of highly creative students. *Journal of Experimental Education,32*, 401-404

APPENDICES

APPENDIX 5

(Appendices for chapter 5)

5.1 Descriptive statistics for raw test scores at each age.

5.2 Test intercorrelations at each age, after correcting for the effects of attenuation due to test unreliability.

5.3 Inter-test correlations at each age, using normalised data.

5.4 Correlations at each age, algebraically corrected for the effects of attenuation due to range restriction.

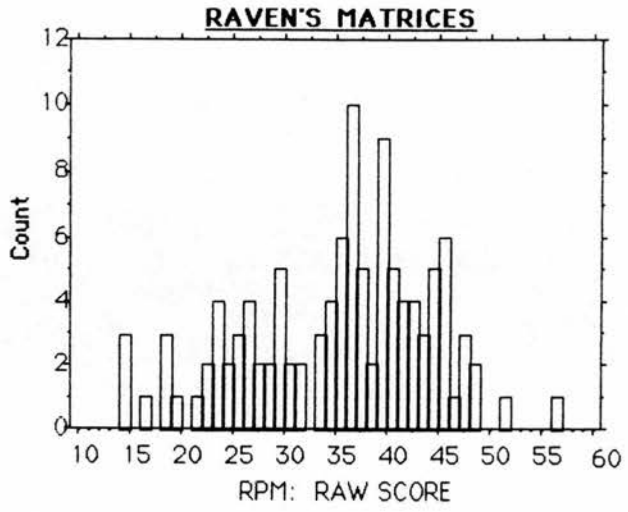
Part 1: Matrices conaining two estimates of unrestricted correlation, each treating one variable in the correlated pair as the selector.
Part 2: Range-corrected correlations, derived from the average of two estimates, each treating one variable in the correlated pair as the selector.

APPENDIX 5.1

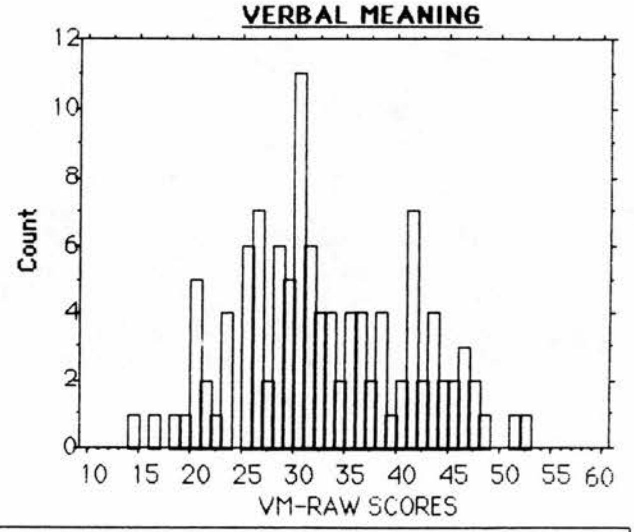
DESCRIPTIVE STATISTICS FOR RAW TEST SCORES AT EACH AGE.

APPENDIX 5.1 (a)

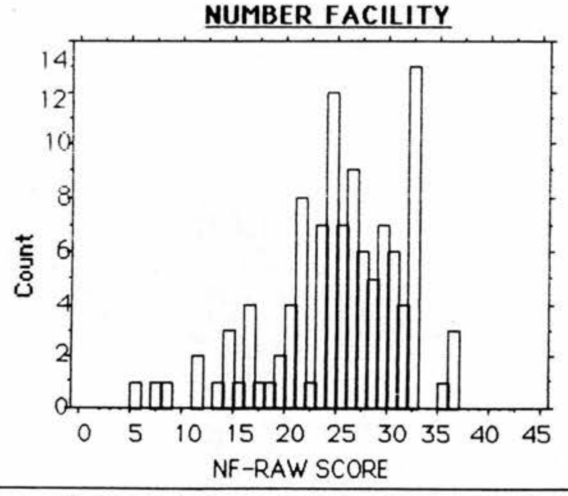
EIGHT YEAR OLDS - RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES



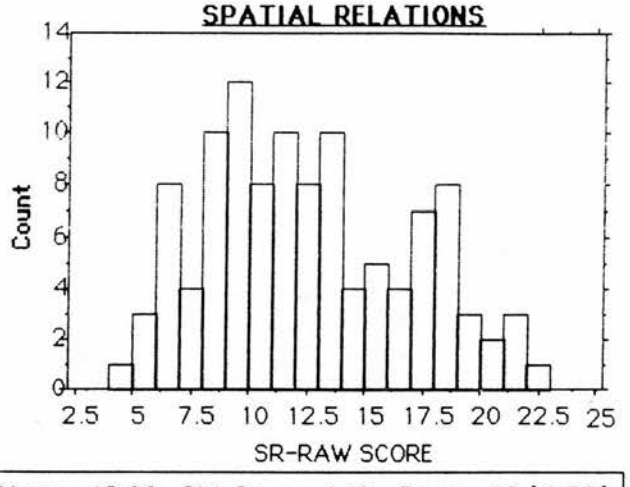
Mean=34.69 Std. Dev. = 8.90 Range= 42 [14-56]
Kurtosis = -.387 Skewness = -.415



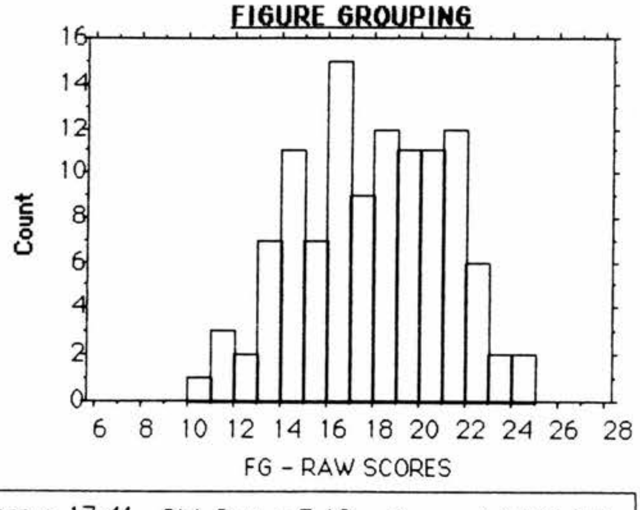
Mean = 32.42 Std. Dev. = 8.26 Range= 38 [14-52]
Kurtosis = -.610 Skewness = .210



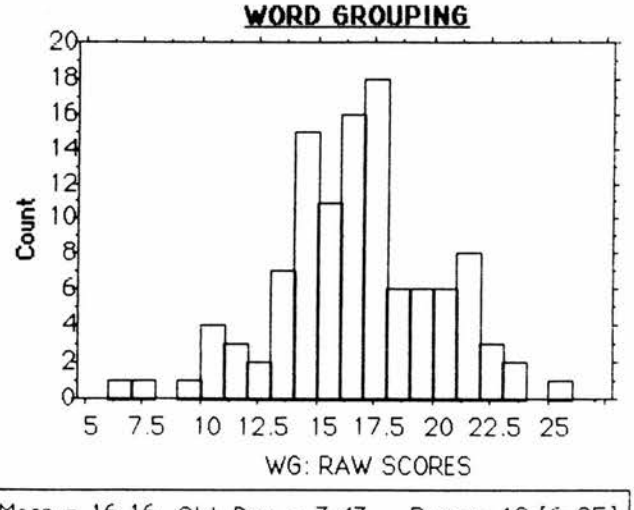
Mean = 24.73 Std. Dev. = 6.306 Range = 31 [5-36]
Kurtosis = .550 Skewness = -.765



Mean = 12.06 Std. Dev. = 4.40 Range= 18 [4-22]
Kurtosis = -.828 Skewness = .330



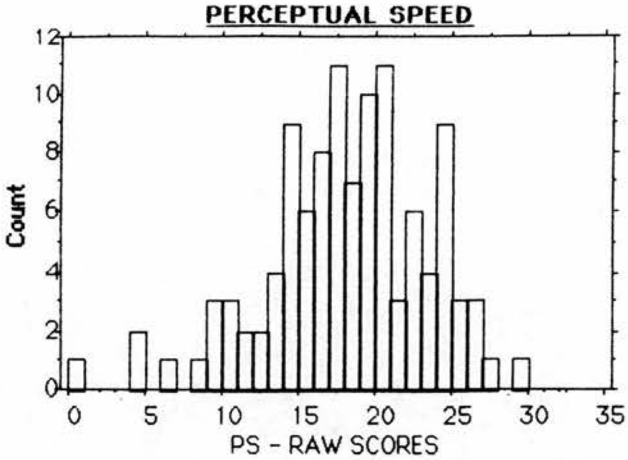
Mean = 17.41 Std. Dev. = 3.18 Range= 14 [10-24]
Kurtosis = -.727 Skewness = -.117



Mean = 16.16 Std. Dev. = 3.43 Range= 19 [6-25]
Kurtosis = .329 Skewness = -.162

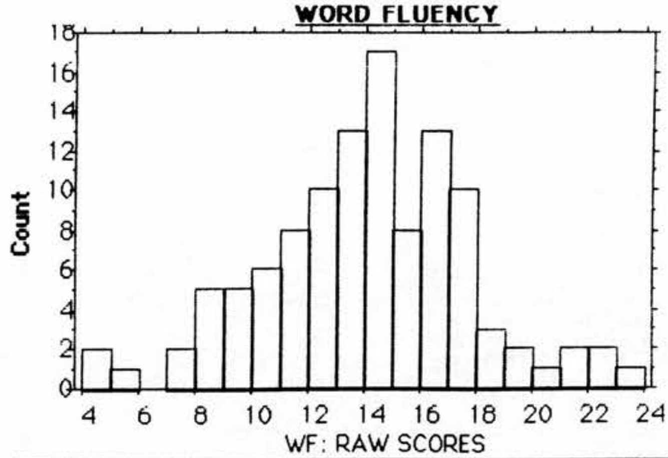
EIGHT YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

PERCEPTUAL SPEED



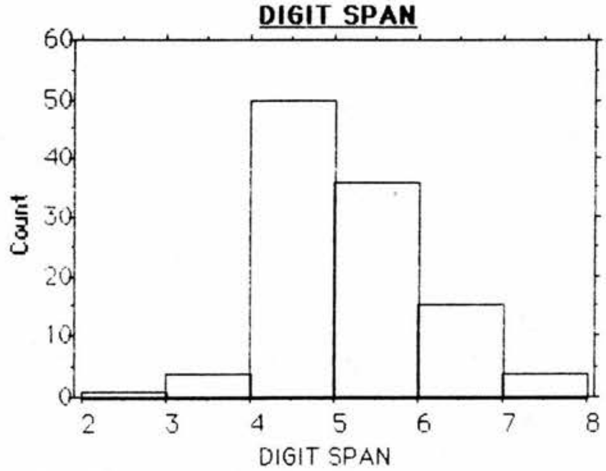
Mean = 17.62 Std. Dev. = 5.25 Range= 29 [0-29]
Kurtosis = .595 Skewness = -.576

WORD FLUENCY



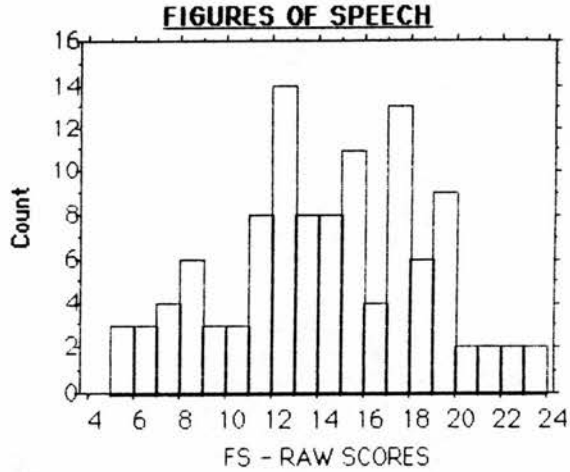
Mean = 13.57 Std. Dev. = 3.66 Range= 19 [4-23]
Kurtosis = .308 Skewness = -.076

DIGIT SPAN



Mean = 4.69 Std. Dev. = .97 Range= 6 [2-8]
Kurtosis = .863 Skewness = .664

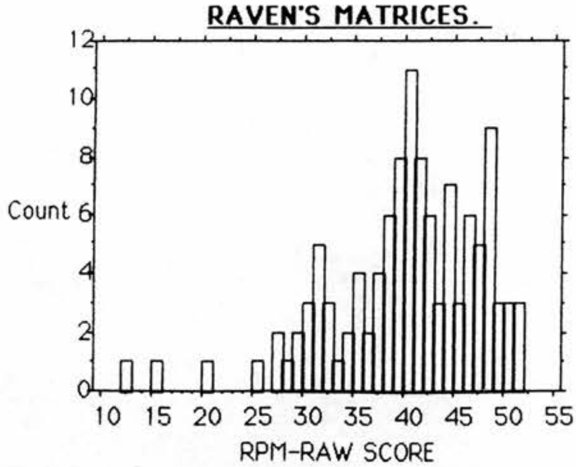
FIGURES OF SPEECH



Mean = 13.87 Std. Dev. = 4.30 Range= 29 [0-29]
Kurtosis = -.597 Skewness = -.112

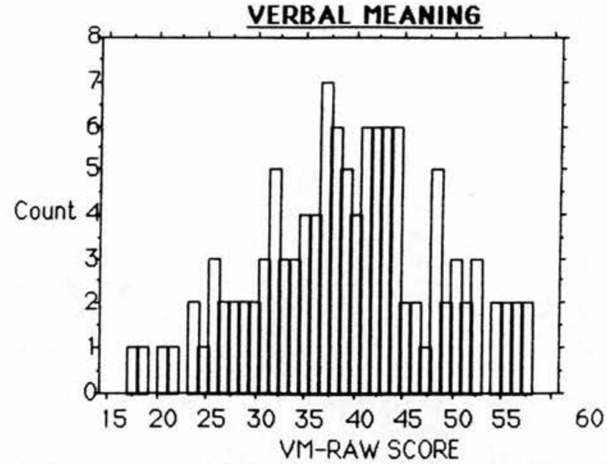
NINE YEAR OLDS – RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

RAVEN'S MATRICES.



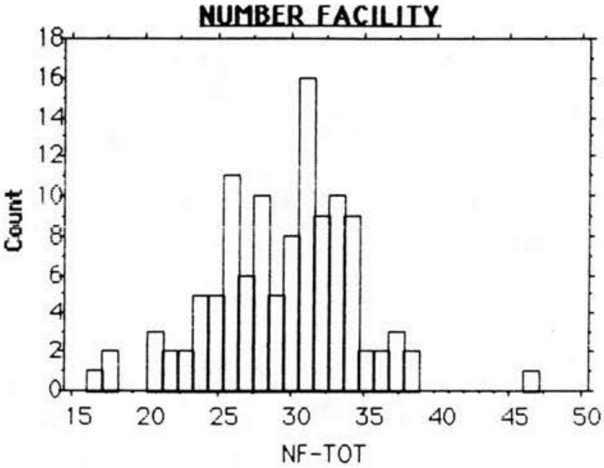
Mean = 39.82 Std. Dev. = 7.38 Range = 39 [12-51]
Kurtosis = 1.581 Skewness = -1.023

VERBAL MEANING



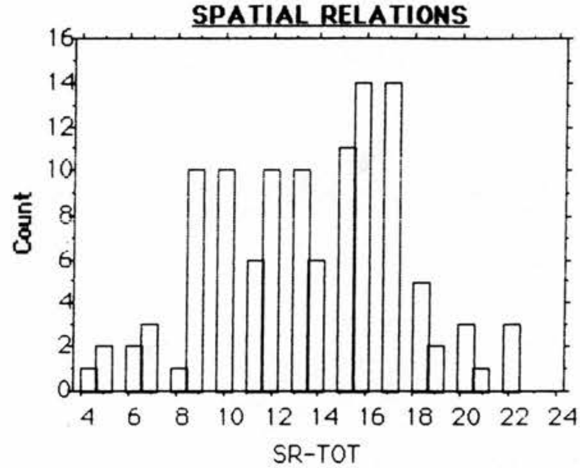
Mean = 39.27 Std. Dev. = 8.83 Range= 40 [17-57]
Kurtosis = -.352 Skewness = -.101

NUMBER FACILITY



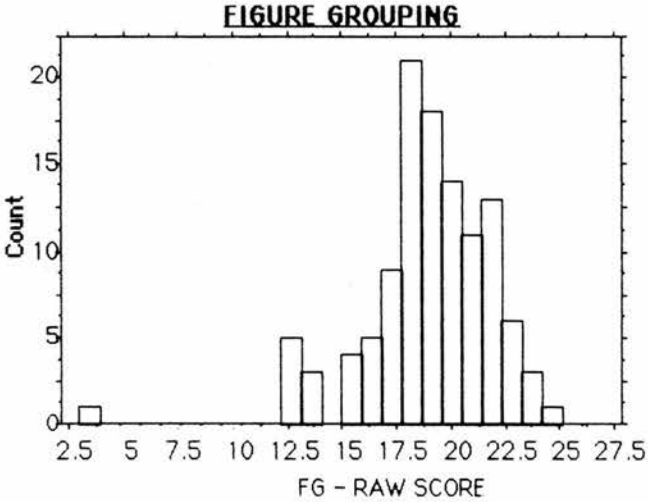
Mean = 29.46 Std. Dev. = 4.65 Range= 30 [16-46]
Kurtosis = .973 Skewness = -.088

SPATIAL RELATIONS



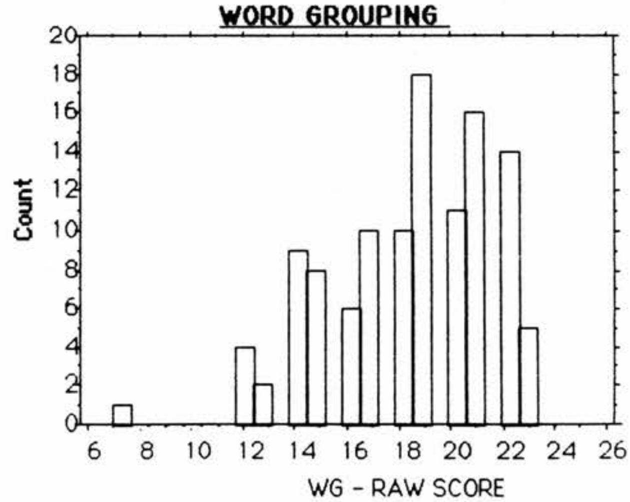
Mean = 13.57 Std. Dev. = 3.92 Range= 40 [17-57]
Kurtosis = -.431 Skewness = -.150

FIGURE GROUPING



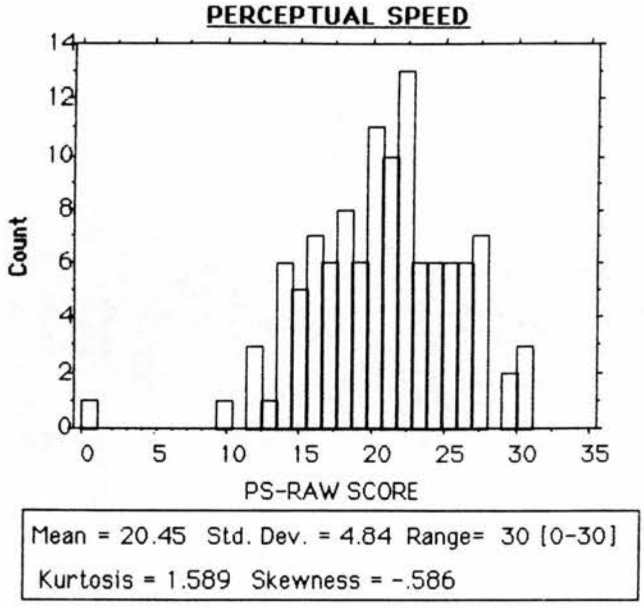
Mean = 18.90 Std. Dev. = 3.06 Range = 22 [3-25]
Kurtosis = 5.120 Skewness = -1.334

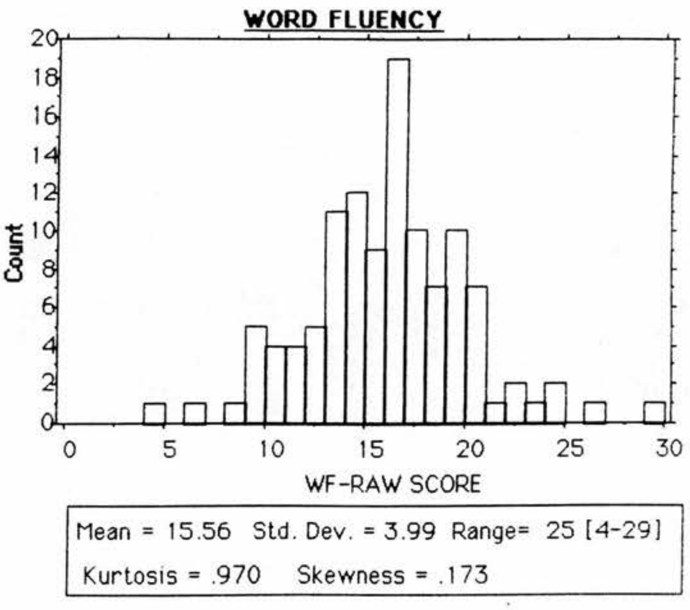
WORD GROUPING

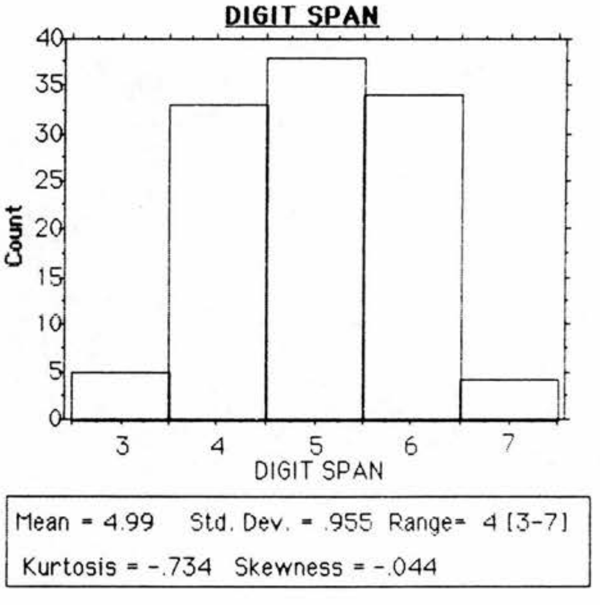


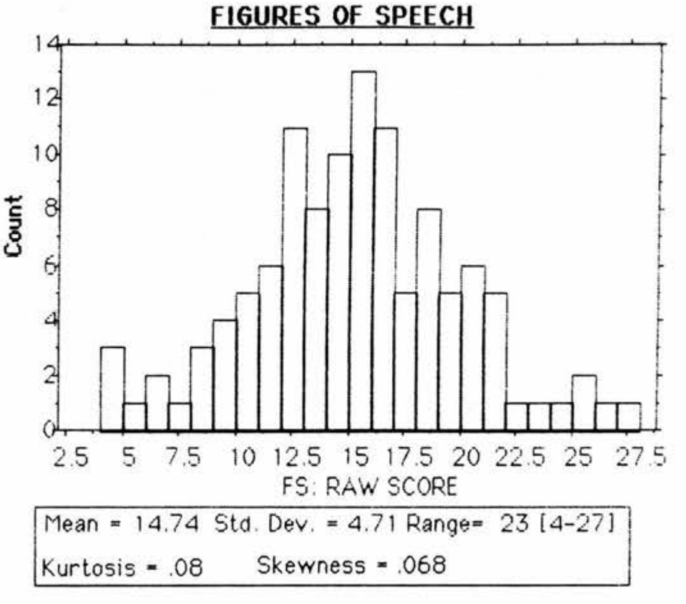
Mean = 18.37 Std. Dev. = 3.13 Range = 16 [7-23]
Kurtosis = .276 Skewness = -.716

NINE YEAR OLDS - RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES





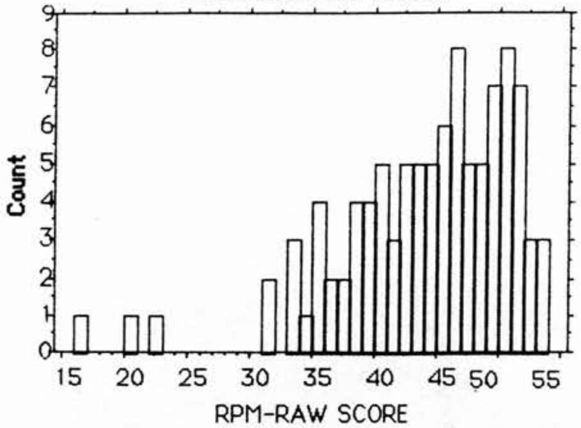




APPENDIX 5.1(2)

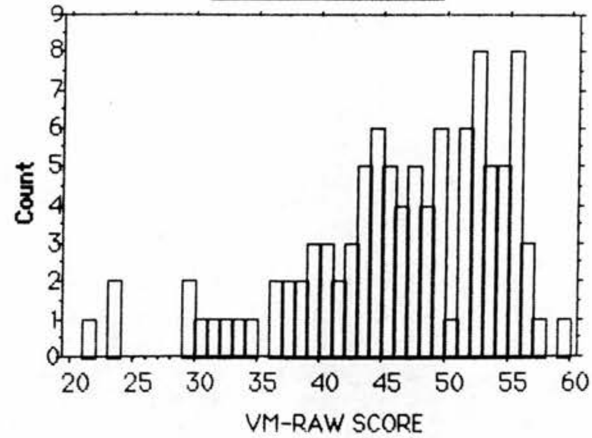
TEN YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

RAVEN'S MATRICES



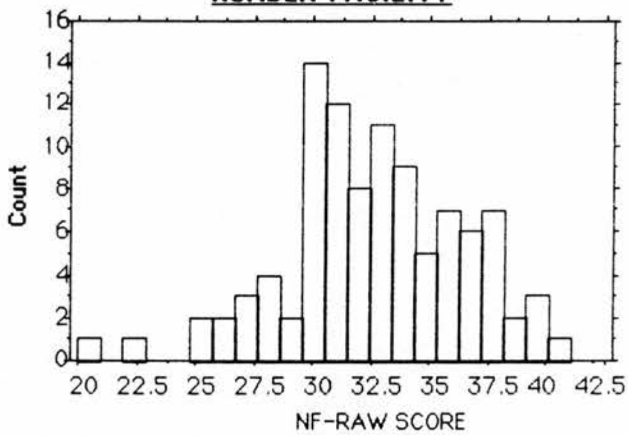
Mean = 43.43 Std. Dev. = 7.07 Range = 37 (16-53)
Kurtosis = 2.299 Skewness = -1.281

VERBAL MEANING



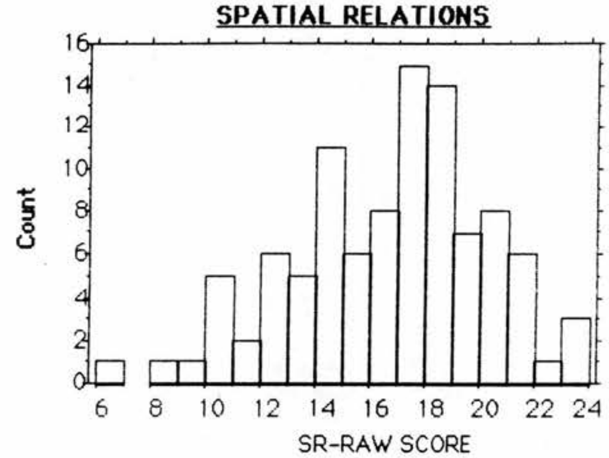
Mean = 45.92 Std. Dev. = 8.10 Range = 38 (21-59)
Kurtosis = .629 Skewness = -.95

NUMBER FACILITY



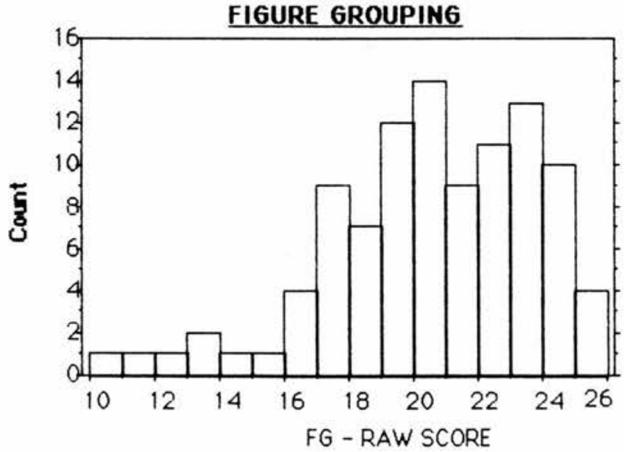
Mean = 32.66 Std. Dev. = 3.99 Range = 21 (20-41)
Kurtosis = .284 Skewness = -.308

SPATIAL RELATIONS



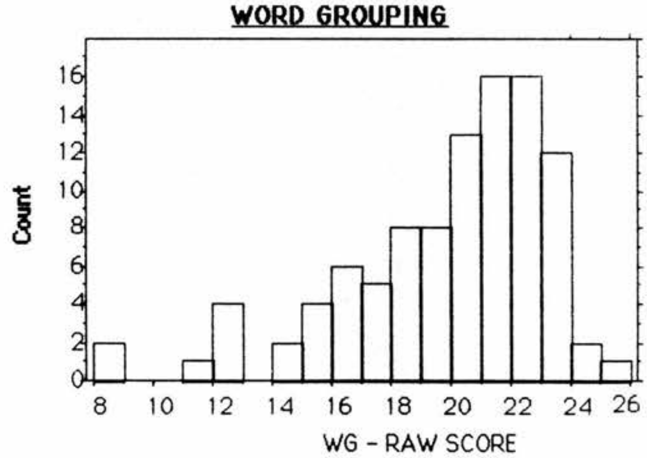
Mean = 16.21 Std. Dev. = 3.52 Range = 17 (6-23)
Kurtosis = -.16 Skewness = -.408

FIGURE GROUPING



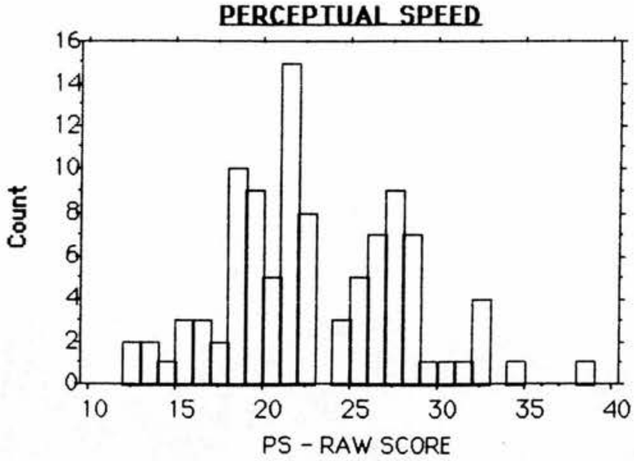
Mean = 20.09 Std. Dev. = 3.21 Range = 15 (10-25)
Kurtosis = .512 Skewness = -.776

WORD GROUPING

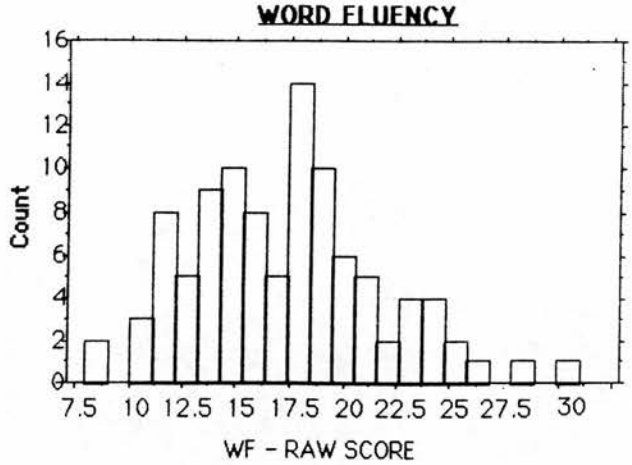


Mean = 19.37 Std. Dev. = 3.46 Range = 17 (8-25)
Kurtosis = 1.221 Skewness = -1.173

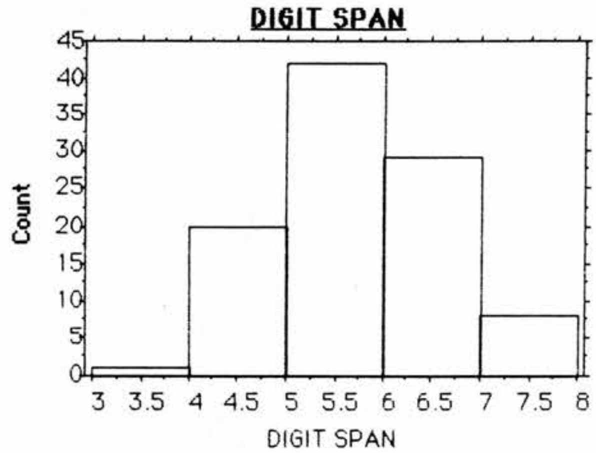
TEN YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES



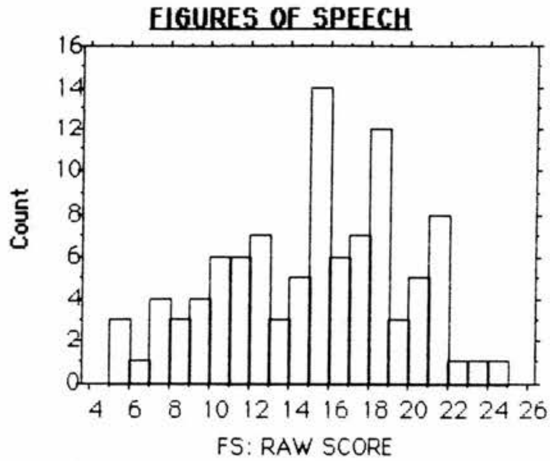
Mean = 22.41 Std. Dev. = 5.15 Range = 26 (12-38)
Kurtosis = -.123 Skewness = .339



Mean = 17.33 Std. Dev. = 4.18 Range = 22 (8-30)
Kurtosis = .096 Skewness = .430



Mean = 5.23 Std. Dev. = .90 Range = 4 (3-7)
Kurtosis = -.492 Skewness = .123

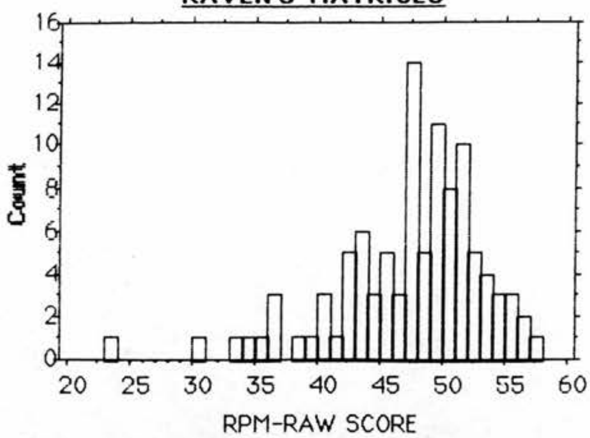


Mean = 14.63 Std. Dev. = 4.54 Range = 19 (5-24)
Kurtosis = -.719 Skewness = -.244

APPENDIX 5·1 (a)

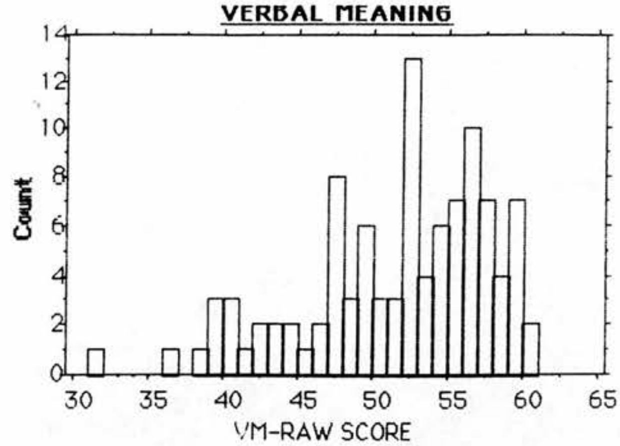
ELEVEN YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

RAVEN'S MATRICES



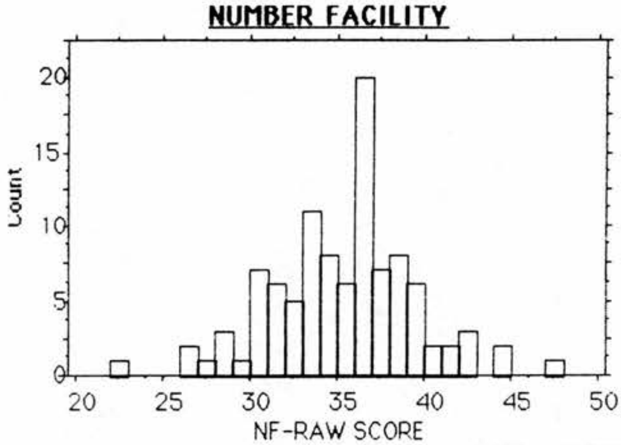
Mean = 46.85 Std. Dev. = 5.87 Range = 34 (23-57)
Kurtosis = 2.13 Skewness = -1.179

VERBAL MEANING



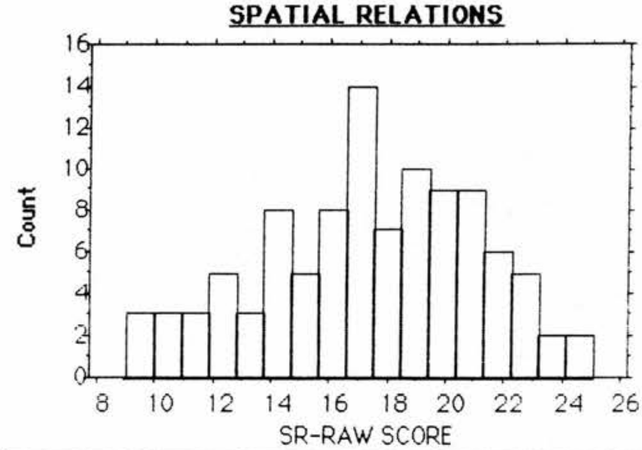
Mean = 51.14 Std. Dev. = 6.17 Range = 29 (31-60)
Kurtosis = .212 Skewness = -.838

NUMBER FACILITY



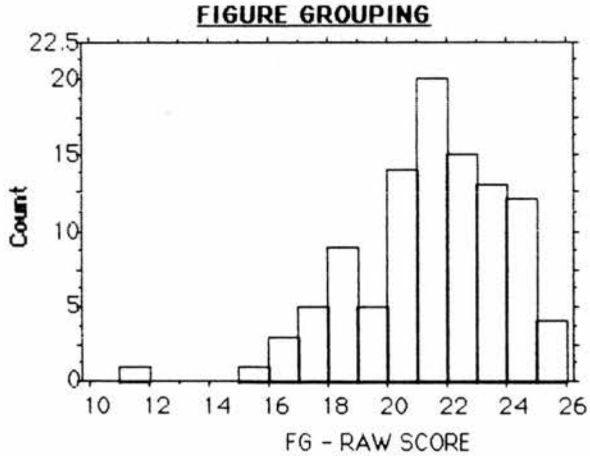
Mean = 34.85 Std. Dev. = 4.14 Range = 25 (22-47)
Kurtosis = .641 Skewness = -.079

SPATIAL RELATIONS



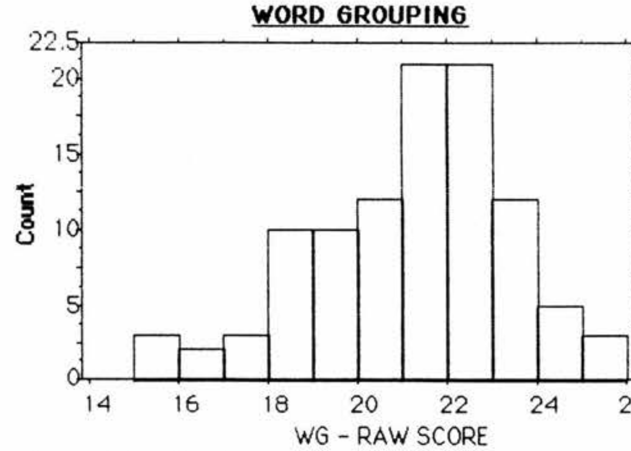
Mean = 17.37 Std. Dev. = 3.90 Range = 16 (9-25)
Kurtosis = -.594 Skewness = -.264

FIGURE GROUPING



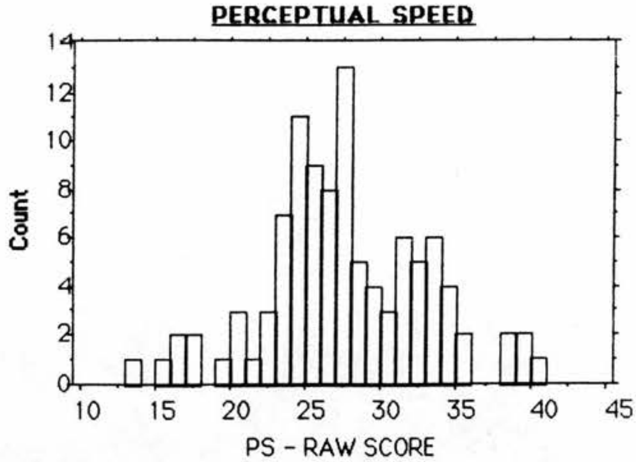
Mean = 20.91 Std. Dev. = 2.53 Range = 14 (11-25)
Kurtosis = 1.176 Skewness = -.827

WORD GROUPING

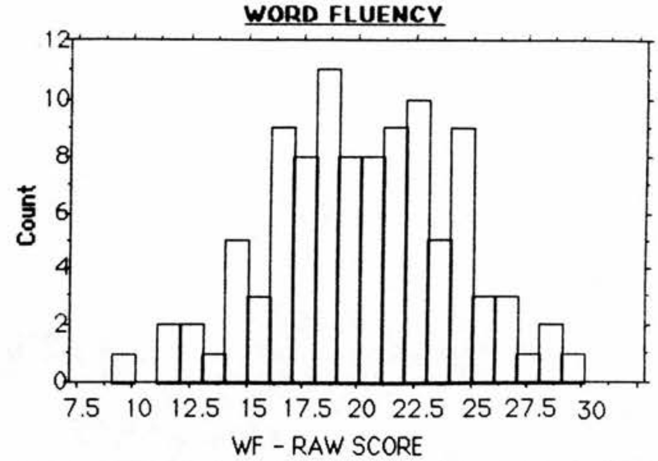


Mean = 20.71 Std. Dev. = 2.23 Range = 10 (15-25)
Kurtosis = -.015 Skewness = -.507

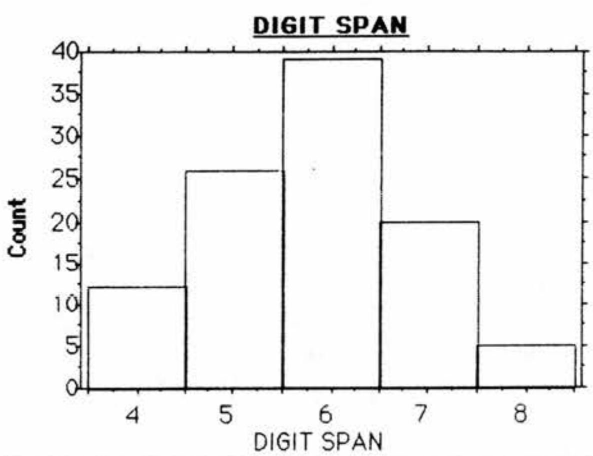
ELEVEN YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES



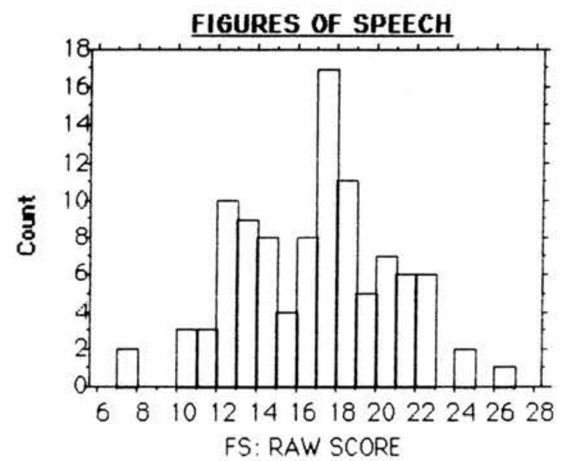
Mean - 27.05 Std. Dev. - 5.33 Range - 27 (13-40)
Kurtosis - .170 Skewness - .041



Mean - 19.69 Std. Dev. - 4.17 Range - 21 (9-30)
Kurtosis - -.185 Skewness - .01



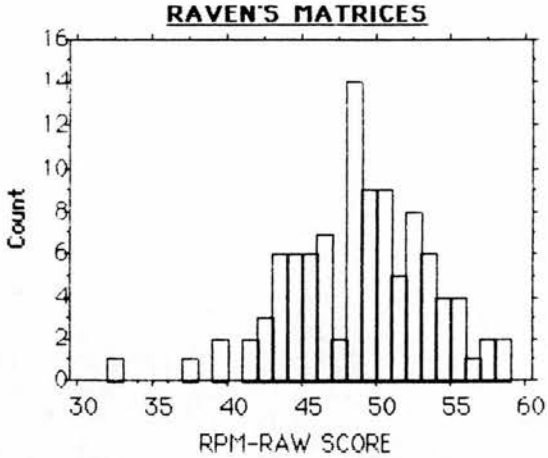
Mean = 5.80 Std. Dev. = 1.04 Range = 4 (4-8)
Kurtosis = -.515 Skewness = .031



Mean = 16.35 Std. Dev. = 3.75 Range = 19 (7-26)
Kurtosis = -.293 Skewness = -.035

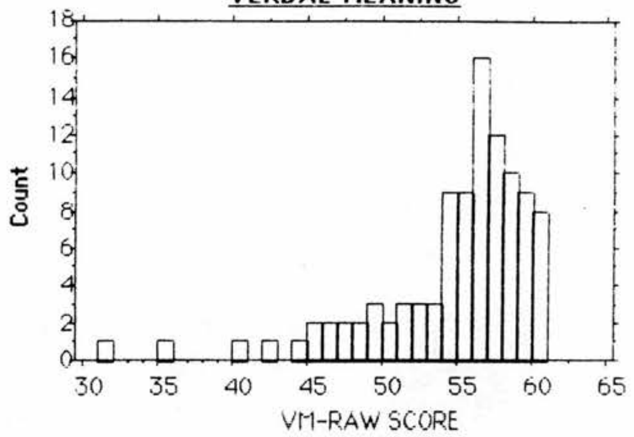
TWELVE YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

RAVEN'S MATRICES



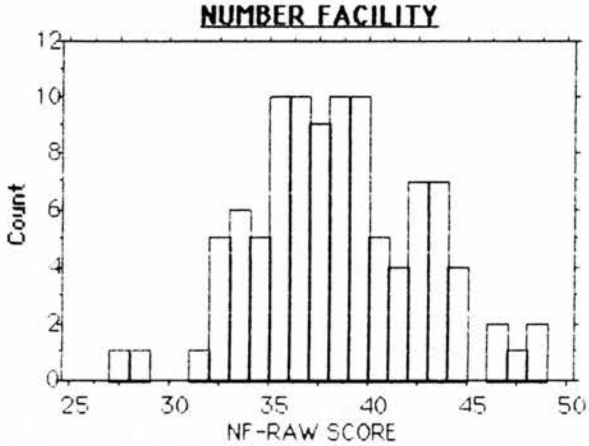
Mean = 48.37 Std. Dev. = 4.71 Range= 26 (32-58)
Kurtosis = .545 Skewness = -.421

VERBAL MEANING



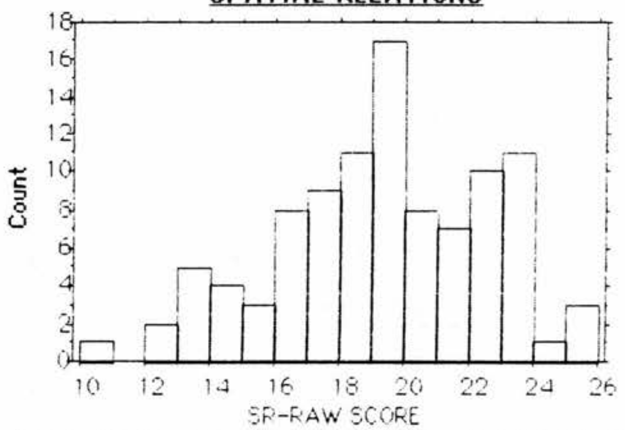
Mean = 54.31 Std. Dev. = 5.30 Range= 29 (31-60)
Kurtosis = 4.264 Skewness = -1.862

NUMBER FACILITY



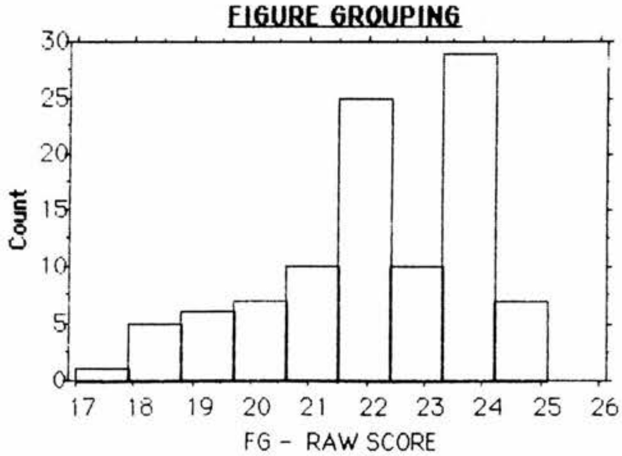
Mean = 37.97 Std. Dev = 4.17 Range= 21 (27-48)
Kurtosis = -.104 Skewness = .149

SPATIAL RELATIONS



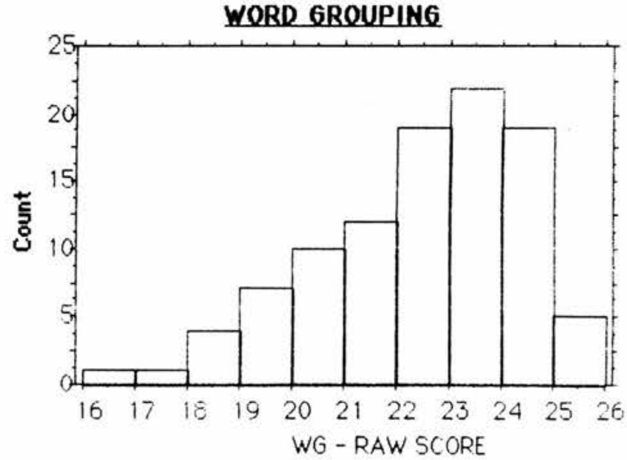
Mean = 18.81 Std. Dev = 3.27 Range = 15 (10-25)
Kurtosis = -.395 Skewness = -.312

FIGURE GROUPING



Mean = 22.22 Std. Dev. = 1.96 Range = 8 (17-25)
Kurtosis = -.33 Skewness = -.665

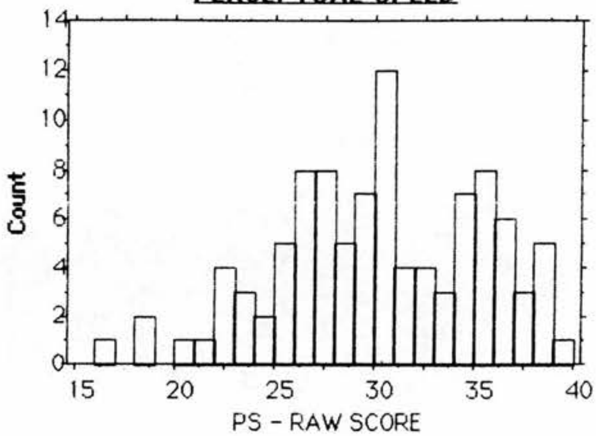
WORD GROUPING



Mean = 21.95 Std Dev. = 1.97 Range = 9 (16-25)
Kurtosis = -.06 Skewness = -.683

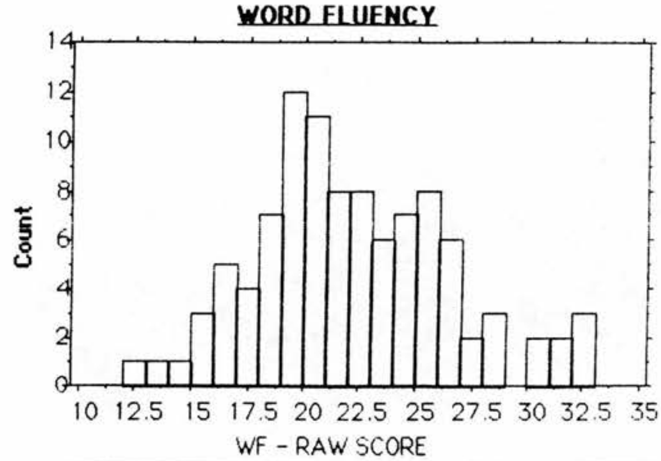
TWELVE YEAR OLDS: RAW SCORE DISTRIBUTIONS AND SUMMARY DESCRIPTIVES

PERCEPTUAL SPEED



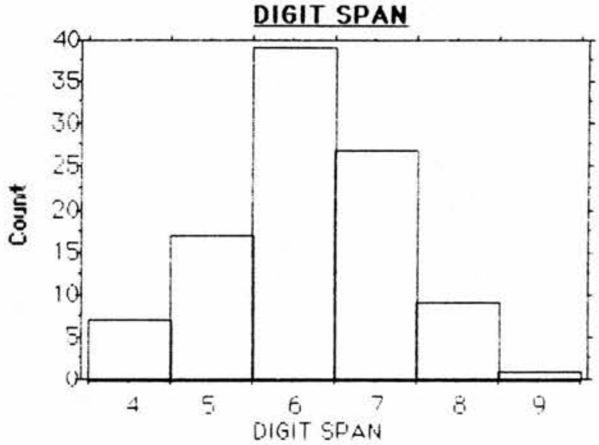
Mean = 29.75 Std. Dev. = 5.13 Range = 23 (16-39)
Kurtosis = -.458 Skewness = -.278

WORD FLUENCY



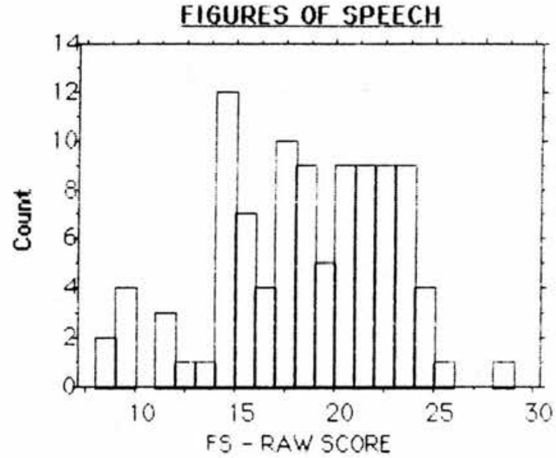
Mean = 21.68 Std. Dev. = 4.36 Range = 20 (12-32)
Kurtosis = -.143 Skewness = .384

DIGIT SPAN



Mean = 6.17 Std. Dev. = 1.07 Range = 5 (4-9)
Kurtosis = -.197 Skewness = -.047

FIGURES OF SPEECH



Mean = 17.97 Std. Dev = 4.26 Range = 20 (8-28)
Kurtosis = -.364 Skewness = -.380

APPENDIX 5.2

TEST INTERCORRELATIONS AT EACH AGE AFTER CORRECTING FOR THE EFFECTS OF ATTENUATION DUE TO UNRELIABILITY.

EIGHT YEAR OLDS

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.386	1								
NF	.510	.601	1							
SR	.614	.483	.675	1						
FG	.590	.330	.564	.601	1					
WG	.585	.713	.713	.653	.533	1				
PS	.429	.447	.593	.601	.396	.554	1			
WF	.362	.570	.644	.436	.302	.597	.454	1		
DS	.109	.078	.164	.068	-.101	.246	.191	.221	1	
FS	.280	.363	.525	.477	.375	.529	.439	.466	.171	1

Average of each test's correlations with all others (excluding diagonals)

Ravens Matrices	.429	Spatial Relations	.512	Word Fluency	.450
Verbal Meaning	.441	Figure Grouping	.399	Digit Span	.128
Number Facility	.554	Word Grouping	.569	Figures of Speech	.403
		Perceptual Speed	.456		

AVERAGE INTER-TEST CORRELATION (excluding diagonals)= .434

NINE YEAR OLDS

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.515	1								
NF	.497	.504	1							
SR	.621	.450	.467	1						
FG	.531	.502	.590	.534	1					
WG	.504	.589	.747	.544	.723	1				
PS	.291	.327	.592	.302	.364	.565	1			
WF	.172	.167	.225	.235	.037	.399	.241	1		
DS	.279	.180	.204	.085	.012	.038	.198	.129	1	
FS	.269	.308	.317	.334	.330	.523	.332	.455	.222	1

Average of each test's correlations with all others (excluding diagonals)

Ravens Matrices	.409	Spatial Relations	.397	Word Fluency	.229
Verbal Meaning	.394	Figure Grouping	.403	Digit Span	.150
Number Facility	.460	Word Grouping	.515	Figures of Speech	.344
		Perceptual Speed	.357		

AVERAGE INTER-TEST CORRELATION (excluding diagonals) = .366

TEN YEAR OLDS

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.656	1								
NF	.512	.429	1							
SR	.620	.516	.518	1						
FG	.646	.502	.576	.658	1					
WG	.668	.621	.504	.538	.659	1				
PS	.336	.195	.564	.523	.386	.417	1			
WF	.174	.384	.411	.303	.240	.128	.326	1		
DS	.130	.156	.119	.133	.061	.099	.105	.071	1	
FS	.129	.107	.318	.166	.021	.044	.172	.363	.204	1

Average of each test's correlations with all others (excluding diagonals)

Ravens Matrices .430
Verbal Meaning .396
Number Facility .439

Spatial Relations .442
Figure Grouping .417
Word Grouping .409
Perceptual Speed .336

Word Fluency .267
Digit Span .120
Figures of Speech .170

AVERAGE INTER-TEST CORRELATION (excluding diagonals) = .342

ELEVEN YEAR OLDS

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.380	1								
NF	.547	.339	1							
SR	.553	.299	.261	1						
FG	.603	.354	.298	.479	1					
WG	.643	.393	.426	.480	.732	1				
PS	.397	.311	.491	.450	.325	.526	1			
WF	.228	.317	.384	.116	.080	.213	.332	1		
DS	.473	.326	.302	.288	.313	.258	.243	-.115	1	
FS	.181	.123	.102	.327	.198	.230	.290	.320	.157	1

Average of each test's correlations with all others (excluding diagonals)

Ravens Matrices .445
Verbal Meaning .316
Number Facility .350

Spatial Relations .361
Figure Grouping .376
Word Grouping .433
Perceptual Speed .374

Word Fluency .208
Digit Span .249
Figures of Speech .214

AVERAGE INTER-TEST CORRELATION (excluding diagonals) = .333

TWELVE YEAR OLDS

	Raven's Matrices	Verbal Meaning	Number Facility	Spatial Relations	Figure Grouping	Word Grouping	Perceptual Speed	Word Fluency	Digit Span	Figures of speech
RPM	1									
VM	.440	1								
NF	.468	.117	1							
SR	.461	.543	.228	1						
FG	.562	.209	.464	.498	1					
WG	.413	.387	.196	.184	.663	1				
PS	.440	.159	.549	.342	.330	.355	1			
WF	.344	.384	.350	.342	.344	.452	.469	1		
DS	.231	.176	.272	.041	.290	.215	.287	.255	1	
FS	.399	.388	.332	.274	.250	.237	.227	.451	.246	1

Average of each test's correlations with all others (excluding diagonals)

Raven's Matrices	.418	Spatial Relations	.324	Word Fluency	.377
Verbal Meaning	.311	Figure Grouping	.401	Digit Span	.224
Number Facility	.331	Word Grouping	.345	Figures of Speech	.312
		Perceptual Speed	.351		

AVERAGE INTER-TEST CORRELATION (excluding diagonals) = .339

APPENDIX 5.3

INTER-TEST CORRELATIONS AT EACH AGE, USING NORMALISED DATA

5.3 (a) EIGHT YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.349	.493	.53	.451	.475	.404	.364	.101	.241
Verbal Meaning	.349	•	.534	.403	.239	.547	.399	.534	.074	.3
Number Facility	.493	.534	•	.607	.433	.582	.494	.626	.155	.451
Spatial Relations	.53	.403	.607	•	.421	.469	.521	.371	.063	.388
Figure Grouping	.451	.239	.433	.421	•	.332	.276	.233	-.072	.245
Word Grouping	.475	.547	.582	.469	.332	•	.408	.476	.194	.363
Perceptual Speed	.404	.399	.494	.521	.276	.408	•	.406	.176	.351
Word Fluency	.364	.534	.626	.371	.233	.476	.406	•	.217	.372
Digit Span	.101	.074	.155	.063	-.072	.194	.176	.217	•	.135
Figures of Speech	.241	.3	.451	.388	.245	.363	.351	.372	.135	•
AVERAGE FOR TEST	.379	.376	.486	.419	.284	.427	.382	.400	.116	.316

5.3 (b) NINE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.465	.486	.534	.491	.397	.280	.152	.208	.253
Verbal Meaning	.465	•	.457	.357	.392	.448	.304	.17	.15	.254
Number Facility	.486	.457	•	.373	.433	.579	.545	.211	.19	.277
Spatial Relations	.534	.357	.373	•	.404	.404	.259	.18	.066	.258
Figure Grouping	.491	.392	.433	.404	•	.417	.274	.04	.001	.246
Word Grouping	.397	.448	.579	.404	.417	•	.413	.298	.016	.364
Perceptual Speed	.280	.304	.545	.259	.274	.413	•	.224	.166	.295
Word Fluency	.152	.17	.211	.18	.04	.298	.224	•	.095	.394
Digit Span	.208	.15	.19	.066	.001	.016	.166	.095	•	.175
Figures of Speech	.253	.254	.277	.258	.246	.364	.295	.394	.175	•
AVERAGE FOR TEST	.363	.333	.395	.315	.300	.371	.307	.196	.119	.280

5.3 (c) TEN YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.53	.429	.44	.478	.525	.285	.128	.112	.095
Verbal Meaning	.53	•	.384	.369	.381	.491	.141	.353	.142	.135
Number Facility	.429	.384	•	.391	.483	.439	.52	.369	.096	.28
Spatial Relations	.44	.369	.391	•	.463	.412	.411	.22	.09	.116
Figure Grouping	.478	.381	.483	.463	•	.487	.311	.201	.008	.005
Word Grouping	.525	.491	.439	.412	.487	•	.413	.133	.145	.018
Perceptual Speed	.285	.141	.52	.411	.311	.413	•	.278	.097	.137
Word Fluency	.128	.353	.369	.22	.201	.133	.278	•	.066	.3
Digit Span	.112	.142	.096	.09	.008	.145	.097	.066	•	.151
Figures of Speech	.095	.135	.28	.116	.005	.018	.137	.3	.151	•
AVERAGE FOR TEST	.336	.325	.377	.324	.313	.340	.288	.228	.101	.137

APPENDIX 5.3 ctd.
INTER-TEST CORRELATIONS AT EACH AGE USING NORMALISED DATA

5.3 (d) ELEVEN YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.355	.454	.446	.348	.448	.367	.151	.416	.093
Verbal Meaning	.355	•	.281	.247	.253	.319	.311	.26	.302	.087
Number Facility	.454	.281	•	.233	.156	.348	.475	.354	.255	.088
Spatial Relations	.446	.247	.233	•	.344	.342	.393	.091	.25	.24
Figure Grouping	.348	.253	.156	.344	•	.445	.261	.037	.23	.134
Word Grouping	.448	.319	.348	.342	.445	•	.399	.149	.208	.129
Perceptual Speed	.367	.311	.475	.393	.261	.399	•	.304	.219	.229
Word Fluency	.151	.26	.354	.091	.037	.149	.304	•	-.105	.268
Digit Span	.416	.302	.255	.25	.23	.208	.219	-.105	•	.132
Figures of Speech	.093	.087	.088	.24	.134	.129	.229	.268	.132	•
AVERAGE FOR TEST	.342	.268	.294	.287	.245	.310	.329	.168	.212	.156

5.3 (e) TWELVE YEAR OLDS

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	•	.371	.415	.379	.385	.264	.417	.31	.19	.334
Verbal Meaning	.371	•	.131	.448	.225	.306	.235	.337	.187	.362
Number Facility	.415	.131	•	.194	.325	.163	.528	.329	.259	.312
Spatial Relations	.379	.448	.194	•	.315	.097	.296	.275	.034	.23
Figure Grouping	.385	.225	.325	.315	•	.329	.25	.224	.168	.168
Word Grouping	.264	.306	.163	.097	.329	•	.268	.316	.136	.164
Perceptual Speed	.417	.235	.528	.296	.25	.268	•	.433	.267	.23
Word Fluency	.31	.337	.329	.275	.224	.316	.433	•	.217	.392
Digit Span	.19	.187	.259	.034	.168	.136	.267	.217	•	.225
Figures of Speech	.334	.362	.312	.23	.168	.164	.23	.392	.225	•
AVERAGE FOR TEST	.341	.289	.295	.252	.265	.227	.325	.315	.187	.268

APPENDIX 5.4
CORRELATIONS AT EACH AGE, ALGEBRAICALLY CORRECTED FOR THE EFFECTS OF ATTENUATION DUE TO RANGE RESTRICTION.

PART 1:
MATRICES CONTAINING TWO ESTIMATES OF UNRESTRICTED CORRELATION, EACH TREATING ONE VARIABLE IN THE CORRELATED PAIR AS THE SELECTOR.

For each column, the common variable is treated as the selector (X) in calculating the 'corrected' intercorrelations. This yields two estimates for each original coefficient: For example, with eight year olds (below) the correlation between Raven's Matrices and Verbal Meaning is adjusted to .443 when RPM is treated as the direct selector (column 1) and to .338 when VM is nominated the selector (column 2).

EIGHT YEAR OLDS - ASYMMETRICAL MATRIX.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.443	.562	.537	.451	.468	.483	.441	.112	.249
Verbal Meaning	.338	*	.647	.415	.248	.557	.492	.642	.078	.315
Number Facility	.458	.654	*	.59	.432	.57	.642	.719	.168	.464
Spatial Relations	.508	.504	.679	*	.422	.48	.604	.484	.064	.388
Figure Grouping	.425	.312	.5	.424	*	.344	.367	.303	-.083	.268
Word Grouping	.443	.653	.66	.482	.344	*	.522	.59	.212	.395
Perceptual Speed	.382	.503	.646	.522	.302	.441	*	.537	.194	.385
Word Fluency	.328	.632	.711	.389	.235	.485	.516	*	.228	.417
Digit Span	.093	.09	.163	.058	-.074	.19	.215	.266	*	.146
Figures of Speech	.227	.383	.526	.381	.261	.385	.451	.507	.159	*

NINE YEAR OLDS- ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.537	.578	.57	.438	.436	.355	.197	.269	.218
Verbal Meaning	.512	*	.583	.42	.414	.504	.393	.191	.174	.25
Number Facility	.501	.531	*	.441	.49	.637	.666	.259	.2	.26
Spatial Relations	.57	.443	.514	*	.41	.437	.342	.249	.077	.251
Figure Grouping	.467	.466	.598	.438	*	.543	.387	.037	.01	.235
Word Grouping	.446	.54	.72	.447	.523	*	.572	.393	.033	.374
Perceptual Speed	.301	.355	.672	.29	.307	.492	*	.278	.194	.273
Word Fluency	.183	.19	.292	.232	.032	.361	.307	*	.13	.385
Digit Span	.265	.183	.238	.076	.009	.031	.228	.138	*	.167
Figures of Speech	.259	.314	.367	.298	.259	.426	.379	.474	.202	*

TEN YEAR OLDS - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.709	.653	.578	.561	.56	.39	.187	.137	.11
Verbal Meaning	.67	*	.576	.497	.442	.527	.235	.41	.166	.092
Number Facility	.534	.499	*	.495	.502	.426	.622	.433	.126	.272
Spatial Relations	.553	.513	.589	*	.493	.389	.511	.279	.121	.122
Figure Grouping	.626	.546	.685	.583	*	.525	.424	.245	.061	.017
Word Grouping	.626	.634	.606	.472	.525	*	.44	.127	.096	.035
Perceptual Speed	.365	.242	.708	.507	.343	.357	*	.352	.113	.149
Word Fluency	.19	.454	.555	.302	.212	.109	.382	*	.076	.312
Digit Span	.133	.179	.166	.125	.05	.078	.119	.072	*	.163
Figures of Speech	.132	.124	.422	.156	.017	.035	.193	.363	.202	*

ELEVEN - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.512	.644	.496	.523	.637	.419	.233	.412	.17
Verbal Meaning	.44	*	.456	.291	.337	.445	.351	.34	.302	.122
Number Facility	.611	.496	*	.26	.292	.486	.547	.418	.287	.104
Spatial Relations	.58	.419	.344	*	.426	.506	.474	.12	.254	.306
Figure Grouping	.568	.439	.35	.391	*	.647	.312	.074	.247	.167
Word Grouping	.605	.486	.487	.397	.571	*	.498	.198	.206	.197
Perceptual Speed	.469	.465	.634	.444	.319	.583	*	.366	.233	.294
Word Fluency	.281	.473	.522	.117	.081	.259	.385	*	-.111	.325
Digit Span	.523	.463	.404	.274	.293	.296	.271	-.123	*	.153
Figures of Speech	.202	.177	.135	.295	.178	.253	.306	.32	.136	*

TWELVE - ASYMMETRICAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.629	.584	.471	.533	.466	.479	.339	.198	.353
Verbal Meaning	.581	*	.164	.552	.219	.447	.184	.384	.153	.348
Number Facility	.638	.213	*	.263	.487	.255	.623	.374	.253	.319
Spatial Relations	.572	.7	.295	*	.462	.211	.363	.322	.034	.231
Figure Grouping	.566	.267	.466	.399	*	.555	.286	.263	.192	.171
Word Grouping	.47	.498	.226	.166	.528	*	.331	.372	.154	.176
Perceptual Speed	.602	.281	.691	.381	.354	.432	*	.486	.262	.215
Word Fluency	.498	.601	.488	.384	.371	.533	.54	*	.235	.427
Digit Span	.331	.293	.367	.045	.298	.259	.323	.255	*	.220
Figures of Speech	.539	.583	.445	.296	.262	.288	.262	.45	.216	*

APPENDIX 5.4

CORRELATIONS AT EACH AGE, ALGEBRAICALLY CORRECTED FOR THE EFFECTS OF ATTENUATION DUE TO RANGE RESTRICTION.

PART 2: RANGE-CORRECTED CORRELATIONS, DERIVED FROM THE AVERAGE OF TWO ESTIMATES, EACH TREATING ONE VARIABLE IN THE CORRELATED PAIR AS THE SELECTOR.

EIGHT YEAR OLDS - FINAL MATRIX.

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.391	.51	.522	.438	.455	.433	.384	.102	.238
Verbal Meaning	.391	*	.65	.459	.28	.605	.498	.637	.084	.349
Number Facility	.51	.65	*	.634	.466	.615	.644	.715	.166	.495
Spatial Relations	.522	.459	.634	*	.423	.481	.563	.436	.061	.384
Figure Grouping	.438	.28	.466	.423	*	.344	.334	.269	-.079	.264
Word Grouping	.455	.605	.615	.481	.344	*	.481	.537	.201	.39
Perceptual Speed	.433	.498	.644	.563	.334	.481	*	.526	.205	.418
Word Fluency	.384	.637	.715	.436	.269	.537	.526	*	.247	.462
Digit Span	.102	.084	.166	.061	-.079	.201	.205	.247	*	.152
Figures of Speech	.238	.349	.495	.384	.264	.39	.418	.462	.152	*
Average for test	.386	.439	.544	.440	.304	.457	.456	.468	.127	.350

NINE YEAR OLDS-FINAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.525	.54	.57	.453	.441	.328	.19	.267	.239
Verbal Meaning	.525	*	.557	.432	.44	.522	.374	.19	.178	.282
Number Facility	.54	.557	*	.477	.544	.678	.669	.276	.219	.314
Spatial Relations	.57	.432	.477	*	.424	.442	.316	.24	.076	.275
Figure Grouping	.453	.44	.544	.424	*	.533	.347	.035	.01	.247
Word Grouping	.441	.522	.678	.442	.533	*	.532	.377	.032	.4
Perceptual Speed	.328	.374	.669	.316	.347	.532	*	.292	.211	.326
Word Fluency	.19	.19	.276	.24	.035	.377	.292	*	.134	.43
Digit Span	.267	.178	.219	.076	.01	.032	.211	.134	*	.185
Figures of Speech	.239	.282	.314	.275	.247	.4	.326	.43	.185	*
Average for test	.395	.389	.475	.361	.337	.440	.377	.240	.146	.300

TEN YEAR OLDS- FINAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.69	.594	.565	.593	.593	.378	.189	.135	.121
Verbal Meaning	.69	*	.538	.505	.494	.58	.238	.432	.172	.108
Number Facility	.594	.538	*	.542	.593	.516	.665	.494	.146	.347
Spatial Relations	.565	.505	.542	*	.538	.431	.509	.29	.123	.139
Figure Grouping	.593	.494	.593	.538	*	.525	.383	.228	.056	.017
Word Grouping	.593	.58	.516	.431	.525	*	.399	.118	.087	.035
Perceptual Speed	.378	.238	.665	.509	.383	.399	*	.367	.116	.171
Word Fluency	.189	.432	.494	.29	.228	.118	.367	*	.074	.337
Digit Span	.135	.172	.146	.123	.056	.087	.116	.074	*	.183
Figures of Speech	.121	.108	.347	.139	.017	.035	.171	.337	.183	*
Average for test	.429	.418	.493	.405	.381	.365	.358	.281	.121	.162

ELEVEN - FINAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.476	.627	.538	.545	.621	.444	.257	.467	.186
Verbal Meaning	.476	*	.476	.355	.388	.465	.408	.407	.383	.15
Number Facility	.627	.476	*	.302	.321	.486	.59	.47	.346	.119
Spatial Relations	.538	.355	.302	*	.409	.451	.459	.118	.264	.301
Figure Grouping	.545	.388	.321	.409	*	.609	.316	.078	.27	.173
Word Grouping	.621	.465	.486	.451	.609	*	.54	.229	.251	.225
Perceptual Speed	.444	.408	.59	.459	.316	.54	*	.376	.252	.3
Word Fluency	.257	.407	.47	.119	.078	.229	.376	*	-.117	.322
Digit Span	.467	.383	.346	.264	.27	.251	.252	-.117	*	.145
Figures of Speech	.186	.15	.119	.301	.173	.225	.3	.322	.145	*
Average for test	.462	.390	.415	.355	.345	.431	.409	.238	.251	.213

TWELVE - FINAL MATRIX

	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.605	.611	.522	.55	.468	.54	.419	.264	.446
Verbal Meaning	.605	*	.189	.626	.243	.472	.232	.492	.223	.466
Number Facility	.611	.188	*	.279	.476	.24	.657	.431	.31	.382
Spatial Relations	.521	.626	.279	*	.431	.189	.372	.353	.039	.264
Figure Grouping	.549	.243	.477	.431	*	.542	.32	.317	.245	.217
Word Grouping	.468	.472	.24	.188	.541	*	.382	.452	.207	.232
Perceptual Speed	.54	.233	.657	.372	.32	.382	*	.513	.293	.239
Word Fluency	.418	.493	.431	.353	.317	.453	.513	*	.245	.438
Digit Span	.264	.223	.31	.04	.245	.206	.293	.245	*	.218
Figures of Speech	.446	.465	.382	.264	.216	.232	.238	.439	.218	*
Average for test	.492	.394	.397	.342	.371	.354	.394	.407	.227	.322

APPENDIX 6

(Appendices for chapter 6)

6.1 Characteristics of test score distributions in 'low' and 'high' ability sub-groups, selected according to IQ on each of the ten tests.

6.2 Distributions of the ten variables in 'low' and 'high' ability groups, selected according to Raven's Matrices IQ.

6.3 Correlations between standardised test scores in samples below and above the mean IQ of each variable.

6.4 Test reliabilities in 'low' and 'high' IQ samples, selected using each test as the criterion.

6.5 Inter-test correlations in samples above and below mean IQ on each variable, corrected for test unreliability.

6.6 Distributions of test variables in sub-groups of 'low', 'mid' and 'high' ability, selected by Raven's Matrices IQ.

6.7 Inter-test correlations in 'low', 'mid' and 'high' ability groups, selected according to IQ on each test, with the relevant selector excluded from the matrix.

6.8 Test reliabilities in 'low', 'mid' and 'high' ability groups, selected according to IQ on each variable.

6.9 Inter-test correlations in 'low', 'mid' and 'high' ability groups, selected according to IQ on each of the ten sub-tests, after correcting for unreliability.

APPENDIX 6.1

CHARACTERISTICS OF TEST SCORE DISTRIBUTIONS IN 'LOW' AND 'HIGH' ABILITY SUB-GROUPS SELECTED ACCORDING TO IQ ON EACH OF TEN TESTS.

(The selector test is shown in bold)

A. SELECTOR = RAVEN'S MATRICES IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=256]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	87.85	8.68	36.60	.13	-.86
VERBAL MEANING	94.76	13.57	65.07	-.32	-.07
NUMBER FACILITY	94.07	13.49	72.32	-.17	.05
SPATIAL RELATIONS	94.08	12.57	60.01	-.39	-.15
FIGURE GROUPING	94.75	13.68	68.69	-.33	-.04
WORD GROUPING	95.04	13.99	70.32	-.10	.20
PERCEPTUAL SPEED	95.84	14.21	70.81	-.37	-.08
WORD FLUENCY	96.94	14.91	75.58	-.20	.05
DIGIT SPAN	98.38	13.24	75.94	-.06	.16
FIGURES OF SPEECH	97.13	13.40	72.44	-.06	-.04

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=255]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	112.17	8.60	37.39	.04	.82
VERBAL MEANING	105.17	14.13	75.45	-.30	-.13
NUMBER FACILITY	105.78	13.44	70.20	-.05	-.12
SPATIAL RELATIONS	105.71	14.47	75.35	-.26	-.20
FIGURE GROUPING	105.14	13.61	70.32	-.33	-.09
WORD GROUPING	104.59	13.27	75.35	.07	-.29
PERCEPTUAL SPEED	104.09	14.00	68.43	-.37	.01
WORD FLUENCY	102.94	13.57	75.48	-.27	.04
DIGIT SPAN	101.81	13.34	67.39	-.20	-.01
FIGURES OF SPEECH	102.98	15.24	71.66	-.47	-.07

APPENDIX 6.1 CTD.

B. SELECTOR = VERBAL MEANING IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=253]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	94.78	14.00	70.97	-.20	.07
VERBAL MEANING	87.72	8.70	37.16	.20	-.86
NUMBER FACILITY	95.50	13.99	67.22	-.31	-.08
SPATIAL RELATIONS	95.43	14.23	72.44	-.21	.13
FIGURE GROUPING	96.50	13.89	72.90	-.20	-.07
WORD GROUPING	94.60	13.83	70.32	-.03	.20
PERCEPTUAL SPEED	96.28	14.47	70.97	-.47	-.02
WORD FLUENCY	96.57	14.38	76.14	.10	.16
DIGIT SPAN	97.91	13.89	75.88	-.05	.25
FIGURES OF SPEECH	97.43	13.79	71.87	-.07	.11

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=258]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	104.91	13.75	70.68	-.24	-.05
VERBAL MEANING	111.98	8.57	36.95	-.02	.78
NUMBER FACILITY	104.25	14.07	66.08	-.35	.04
SPATIAL RELATIONS	104.36	13.88	70.18	-.20	-.11
FIGURE GROUPING	103.66	14.17	75.48	-.34	-.10
WORD GROUPING	105.16	13.31	75.35	.05	-.24
PERCEPTUAL SPEED	103.45	14.12	75.35	-.14	.04
WORD FLUENCY	103.61	14.26	75.35	-.26	-.17
DIGIT SPAN	102.45	12.44	61.28	-.20	-.07
FIGURES OF SPEECH	102.57	14.99	72.32	-.34	-.41

APPENDIX 6.1 CTD.

C. SELECTOR = NUMBER FACILITY IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=257]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	94.00	13.45	72.32	.15	.16
VERBAL MEANING	95.75	14.43	73.05	-.32	.10
NUMBER FACILITY	87.91	8.67	37.40	.25	-.88
SPATIAL RELATIONS	95.40	14.32	72.44	-.05	.16
FIGURE GROUPING	95.67	14.64	72.90	-.39	.03
WORD GROUPING	94.43	13.33	68.06	-.08	.05
PERCEPTUAL SPEED	95.25	13.99	70.97	.13	.11
WORD FLUENCY	95.54	13.57	69.63	-.36	-.05
DIGIT SPAN	97.89	13.11	69.23	-.15	-.01
FIGURES OF SPEECH	97.00	13.84	72.44	.05	.09

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=262]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	105.67	13.60	70.68	-.13	-.19
VERBAL MEANING	104.20	13.69	75.45	-.01	-.12
NUMBER FACILITY	111.84	8.63	36.95	.12	.86
SPATIAL RELATIONS	104.35	13.55	70.06	-.29	-.09
FIGURE GROUPING	103.99	12.98	70.32	-.30	.06
WORD GROUPING	105.45	13.51	72.19	-.13	-.20
PERCEPTUAL SPEED	106.52	13.15	68.43	-.21	-.11
WORD FLUENCY	104.63	14.28	70.20	-.28	-.02
DIGIT SPAN	102.07	13.59	67.92	-.31	.11
FIGURES OF SPEECH	102.87	15.00	71.77	-.47	-.12

APPENDIX 6.1 CTD.

D. SELECTOR = SPATIAL RELATIONS IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=252]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	94.91	13.99	64.29	-.53	-.09
VERBAL MEANING	95.32	14.85	76.01	.04	.27
NUMBER FACILITY	95.83	14.17	75.58	-.23	-.02
SPATIAL RELATIONS	87.72	85.56	36.56	.16	-.85
FIGURE GROUPING	95.87	13.70	72.90	.04	.06
WORD GROUPING	95.99	14.15	70.32	-.09	.10
PERCEPTUAL SPEED	95.47	13.92	70.68	-.28	.03
WORD FLUENCY	97.42	14.35	70.32	-.31	-.04
DIGIT SPAN	98.99	14.09	75.88	-.09	.02
FIGURES OF SPEECH	97.19	15.27	72.44	-.22	.19

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=269]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	104.99	13.65	70.68	-.31	.09
VERBAL MEANING	104.30	13.24	72.49	-.26	-.18
NUMBER FACILITY	103.99	14.06	70.20	-.30	-.01
SPATIAL RELATIONS	111.47	8.72	37.38	-.03	.79
FIGURE GROUPING	103.77	14.23	70.32	-.32	-.23
WORD GROUPING	103.69	13.80	75.35	-.30	-.14
PERCEPTUAL SPEED	104.32	14.00	72.19	-.28	-.01
WORD FLUENCY	102.35	14.60	72.90	-.20	.01
DIGIT SPAN	101.10	12.84	61.28	-.42	.15
FIGURES OF SPEECH	102.53	13.64	71.66	-.28	-.05

APPENDIX 6.1 CTD.

E. SELECTOR = FIGURE GROUPING IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=268]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	95.02	14.03	70.20	-.30	.05
VERBAL MEANING	96.97	13.74	75.50	-.04	-.03
NUMBER FACILITY	96.46	13.65	68.69	-.35	-.05
SPATIAL RELATIONS	95.41	14.09	70.20	-.28	.05
FIGURE GROUPING	88.52	8.80	36.97	.16	-.87
WORD GROUPING	95.40	13.87	70.32	-.19	.05
PERCEPTUAL SPEED	97.30	14.54	68.55	-.54	-.06
WORD FLUENCY	98.85	14.49	74.92	-.22	-.07
DIGIT SPAN	99.86	13.75	75.88	-.27	.16
FIGURES OF SPEECH	98.01	14.16	71.87	-.11	.16

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=241]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	105.82	13.55	70.68	-.17	-.06
VERBAL MEANING	103.55	15.07	75.45	-.39	-.18
NUMBER FACILITY	104.21	14.56	76.01	-.07	-.08
SPATIAL RELATIONS	105.26	13.49	70.06	-.32	.01
FIGURE GROUPING	112.62	8.00	36.22	.13	.84
WORD GROUPING	105.38	13.29	70.68	-.24	-.10
PERCEPTUAL SPEED	103.94	13.77	75.35	-.04	.12
WORD FLUENCY	101.45	14.69	76.14	-.23	.05
DIGIT SPAN	100.36	13.16	68.70	-.03	-.12
FIGURES OF SPEECH	102.13	14.66	71.66	-.29	-.22

F. SELECTOR = WORD GROUPING IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=258]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	94.91	13.22	70.97	-.30	-.05
VERBAL MEANING	94.63	13.25	67.89	-.16	-.04
NUMBER FACILITY	94.78	13.19	64.17	-.30	-.04
SPATIAL RELATIONS	95.75	13.60	69.53	-.37	.18
FIGURE GROUPING	95.47	14.14	72.90	-.32	-.03
WORD GROUPING	88.12	8.75	37.23	.06	-.80
PERCEPTUAL SPEED	95.05	14.34	70.81	-.47	-.02
WORD FLUENCY	96.97	14.31	75.48	-.10	-.06
DIGIT SPAN	98.68	13.03	74.82	-.04	.17
FIGURES OF SPEECH	97.23	13.75	71.77	.03	.14

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=248]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	105.67	13.68	65.94	-.46	.02
VERBAL MEANING	106.04	13.65	70.70	-.36	-.09
NUMBER FACILITY	105.86	13.79	75.48	.05	-.15
SPATIAL RELATIONS	104.41	14.52	76.01	.16	-.28
FIGURE GROUPING	104.40	13.44	67.89	-.37	.01
WORD GROUPING	112.24	8.24	37.28	.10	.80
PERCEPTUAL SPEED	105.45	13.27	65.94	-.28	.13
WORD FLUENCY	103.03	14.31	70.20	-.47	.01
DIGIT SPAN	102.11	13.62	75.88	-.17	-.08
FIGURES OF SPEECH	102.48	15.05	72.32	-.43	-.21

APPENDIX 6.1 CTD.

G. SELECTOR = PERCEPTUAL SPEED IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=256]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	96.15	13.78	70.41	-.21	.03
VERBAL MEANING	97.13	14.47	75.48	-.10	.05
NUMBER FACILITY	94.21	13.99	70.81	-.04	.15
SPATIAL RELATIONS	95.90	13.94	67.75	-.47	-.02
FIGURE GROUPING	96.84	14.31	72.90	-.09	.01
WORD GROUPING	95.65	13.91	70.32	-.07	.08
PERCEPTUAL SPEED	87.86	8.65	36.01	.18	-.88
WORD FLUENCY	95.79	14.77	75.02	-.18	.18
DIGIT SPAN	98.65	13.49	75.88	-.05	.04
FIGURES OF SPEECH	97.60	14.20	71.87	.10	.17

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=264]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	103.67	14.63	76.01	-.29	-.09
VERBAL MEANING	102.83	14.35	75.45	-.33	-.12
NUMBER FACILITY	105.64	13.20	70.20	-.21	-.07
SPATIAL RELATIONS	103.94	14.18	76.01	-.12	-.04
FIGURE GROUPING	102.92	13.97	70.32	-.40	-.10
WORD GROUPING	104.43	13.62	72.19	-.30	-.15
PERCEPTUAL SPEED	111.75	8.73	36.95	.10	.83
WORD FLUENCY	104.08	13.56	76.14	-.12	-.05
DIGIT SPAN	101.29	13.31	67.39	-.34	.07
FIGURES OF SPEECH	102.75	14.46	72.44	-.38	-.16

APPENDIX 6.1 CTD.

H. SELECTOR = WORD FLUENCY IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=250]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	97.30	14.84	75.58	-.26	-.02
VERBAL MEANING	95.76	14.91	73.05	-.31	.08
NUMBER FACILITY	95.22	14.30	68.55	-.40	.05
SPATIAL RELATIONS	96.88	14.61	72.30	-.22	.11
FIGURE GROUPING	98.34	14.00	68.55	-.25	-.18
WORD GROUPING	96.27	14.31	70.18	-.38	.12
PERCEPTUAL SPEED	95.35	14.15	75.60	-.02	.14
WORD FLUENCY	87.60	8.55	37.14	.22	-.89
DIGIT SPAN	99.22	13.43	75.35	-.11	-.43
FIGURES OF SPEECH	96.18	13.98	71.76	-.14	.25

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=269]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	102.44	14.26	70.68	-.38	.05
VERBAL MEANING	103.93	13.34	70.83	-.31	.05
NUMBER FACILITY	104.33	13.74	76.14	-.13	.02
SPATIAL RELATIONS	102.76	14.20	76.01	-.26	-.10
FIGURE GROUPING	101.31	14.77	76.14	-.39	.02
WORD GROUPING	103.27	13.88	76.01	.02	-.15
PERCEPTUAL SPEED	104.32	14.05	75.45	-.24	-.14
WORD FLUENCY	111.54	8.81	36.97	.15	.87
DIGIT SPAN	101.08	13.47	67.92	-.32	.14
FIGURES OF SPEECH	103.72	14.15	72.44	-.00	-.20

APPENDIX 6.1 CTD.

I. SELECTOR = DIGIT SPAN IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=257]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	97.22	14.81	76.01	-.11	.14
VERBAL MEANING	97.69	13.79	70.68	-.22	-.00
NUMBER FACILITY	97.72	14.51	75.35	-.25	.01
SPATIAL RELATIONS	99.41	14.46	76.01	-.09	-.07
FIGURE GROUPING	98.90	15.22	72.34	-.49	.03
WORD GROUPING	98.17	14.06	70.18	-.24	-.03
PERCEPTUAL SPEED	97.87	14.34	70.97	-.39	-.10
WORD FLUENCY	98.76	14.53	75.58	-.09	-.10
DIGIT SPAN	89.05	7.74	35.50	.56	-.94
FIGURES OF SPEECH	97.82	14.50	71.87	-.16	.11

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=232]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	102.78	13.98	74.92	-.18	-.15
VERBAL MEANING	102.24	15.09	76.01	-.25	-.18
NUMBER FACILITY	102.46	14.12	69.63	-.36	-.03
SPATIAL RELATIONS	100.55	14.61	71.77	-.46	-.01
FIGURE GROUPING	100.55	12.89	76.01	.11	-.05
WORD GROUPING	101.68	14.49	72.19	-.37	-.05
PERCEPTUAL SPEED	102.12	14.72	75.45	-.17	.10
WORD FLUENCY	101.57	14.50	72.77	-.35	.06
DIGIT SPAN	112.31	7.81	35.55	.46	.86
FIGURES OF SPEECH	102.16	14.55	71.77	-.34	-.10

APPENDIX 6.1 CTD.

J. SELECTOR = FIGURES OF SPEECH IQ

"LOW" RANGE ON SELECTOR = 62-99.5, MEAN = 88 [N=248]

	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	97.56	14.54	75.58	-.08	.00
VERBAL MEANING	97.61	14.78	75.88	-.28	.07
NUMBER FACILITY	96.33	14.28	76.14	-.15	.01
SPATIAL RELATIONS	96.72	15.10	72.44	-.24	.04
FIGURE GROUPING	97.76	14.72	76.14	-.27	.11
WORD GROUPING	96.95	14.56	70.32	-.41	.09
PERCEPTUAL SPEED	95.89	14.17	76.01	-.32	-.27
WORD FLUENCY	95.46	14.39	74.92	-.39	.07
DIGIT SPAN	98.07	13.18	75.35	-.03	.10
FIGURES OF SPEECH	87.57	8.48	33.80	-.08	-.79

"HIGH" RANGE ON SELECTOR = 100.5-139, MEAN =112 [N=257]

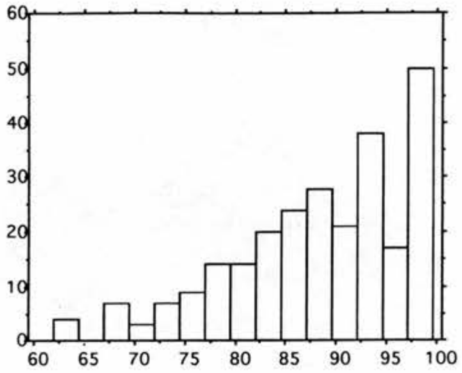
	MEAN	SD	RANGE	KURTOSIS	SKEWNESS
RAVEN'S MATRICES	102.31	14.46	70.68	-.38	-.03
VERBAL MEANING	102.42	14.21	75.48	-.14	-.14
NUMBER FACILITY	102.99	14.24	70.25	-.25	-.02
SPATIAL RELATIONS	103.22	13.63	67.75	-.56	.09
FIGURE GROUPING	101.96	13.99	72.24	-.17	-.18
WORD GROUPING	102.90	13.95	75.35	-.04	-.22
PERCEPTUAL SPEED	103.43	14.37	74.92	-.34	-.01
WORD FLUENCY	104.43	13.77	76.14	-.00	-.01
DIGIT SPAN	102.10	13.43	61.28	-.37	-.05
FIGURES OF SPEECH	112.03	8.66	36.92	.19	.85

APPENDIX 6.2 DISTRIBUTIONS OF ALL TEN VARIABLES IN 'LOW' AND 'HIGH' ABILITY GROUPS SELECTED ACCORDING TO RAVEN'S MATRICES IQ.

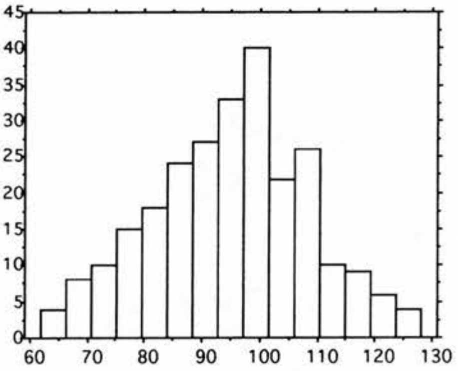
["LOW" = RANGE 66.0 -> 99.5, MEAN 88.0, "HIGH" = RANGE 100.5 - 138.0, MEAN 112.0]

"LOW"

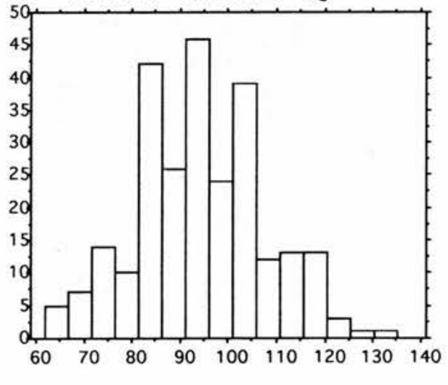
RAVEN'S MATRICES IQ



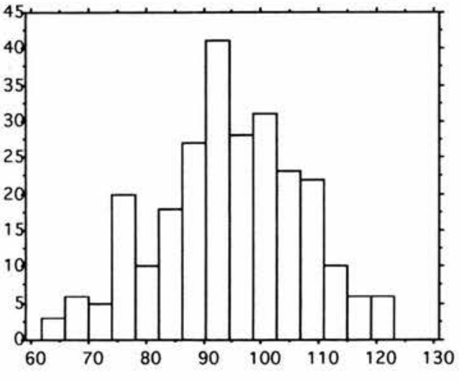
VERBAL MEANING IQ



NUMBER FACILITY IQ

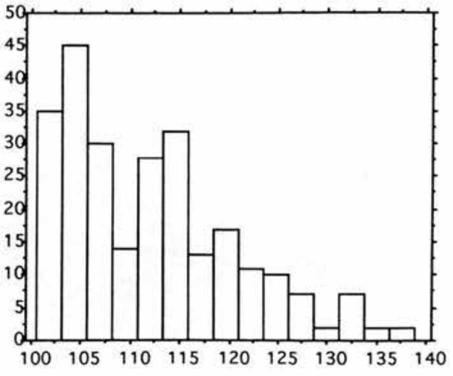


SPATIAL RELATIONS IQ

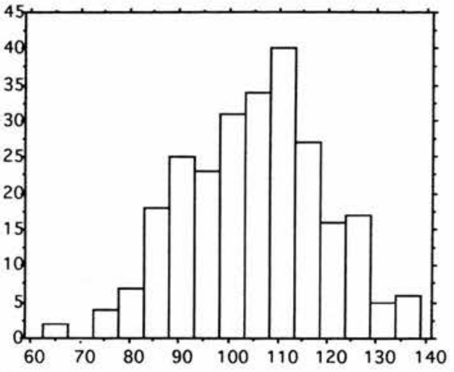


"HIGH"

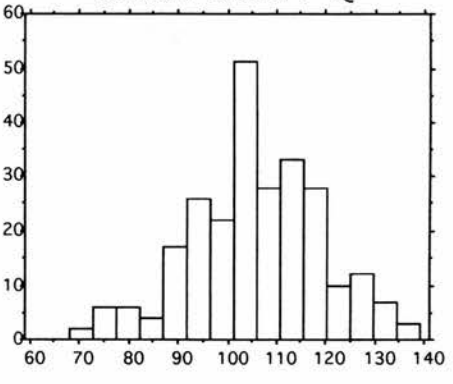
RAVEN'S MATRICES IQ



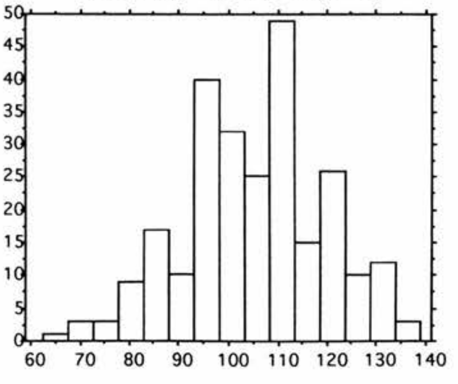
VERBAL MEANING IQ



NUMBER FACILITY IQ

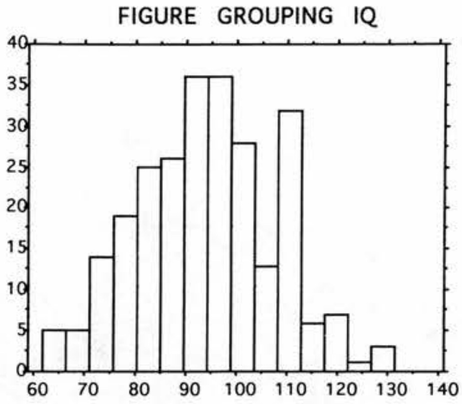


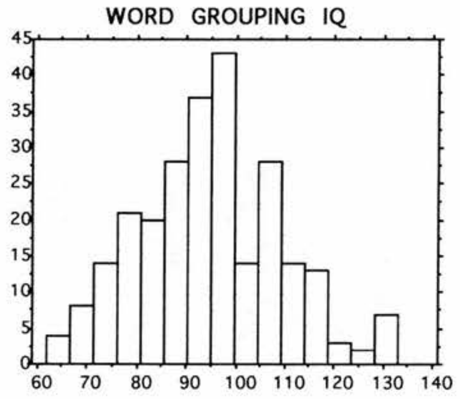
SPATIAL RELATIONS IQ

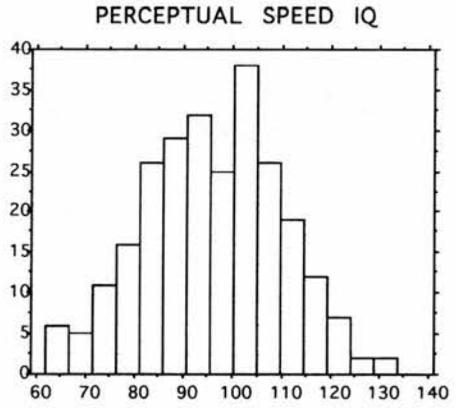


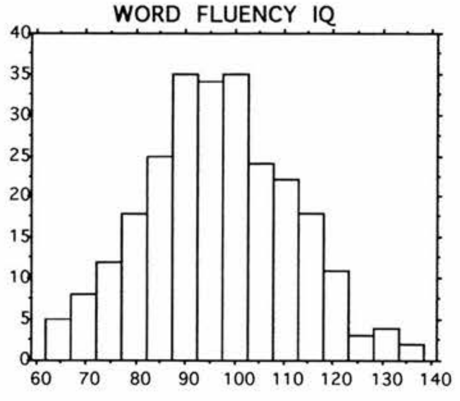
RAVEN'S MATRICES IQ = SELECTOR VARIABLE

"LOW"

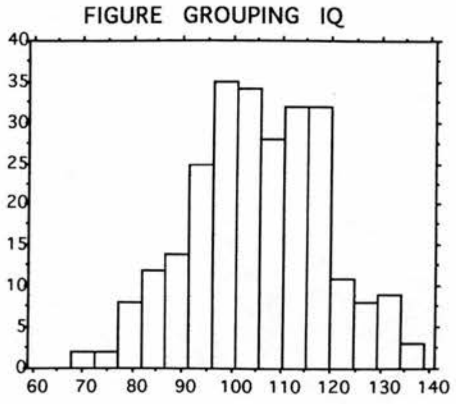


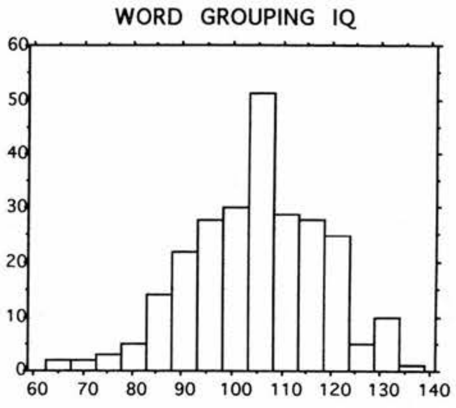


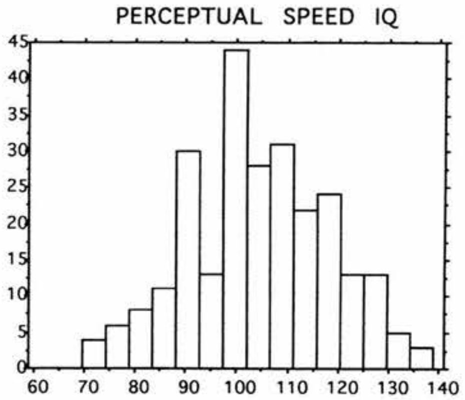


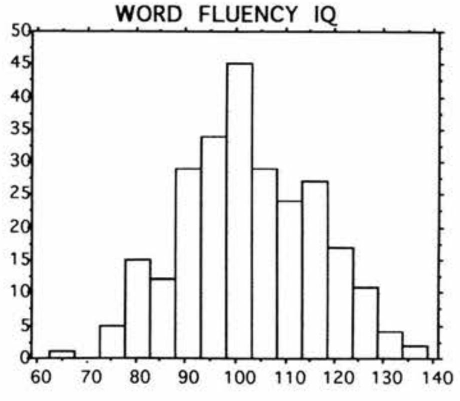


"HIGH"



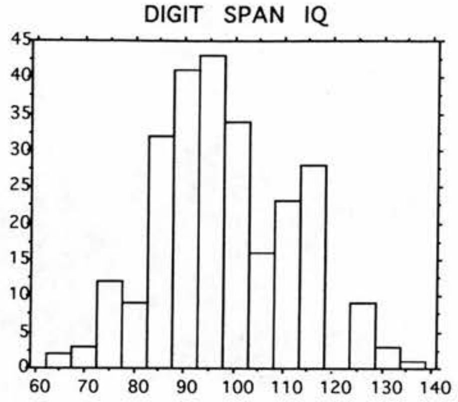


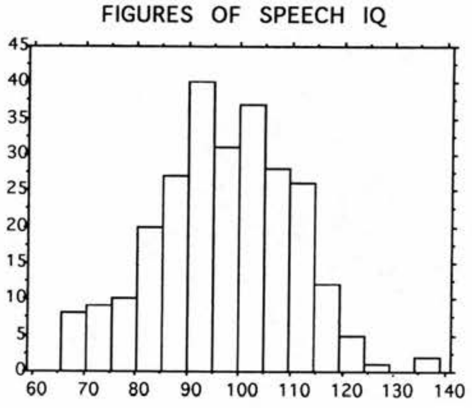




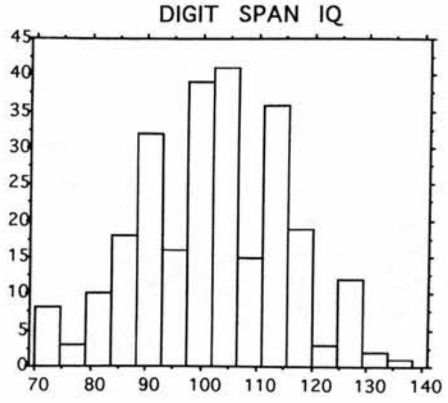
RAVEN'S MATRICES IQ = SELECTOR VARIABLE

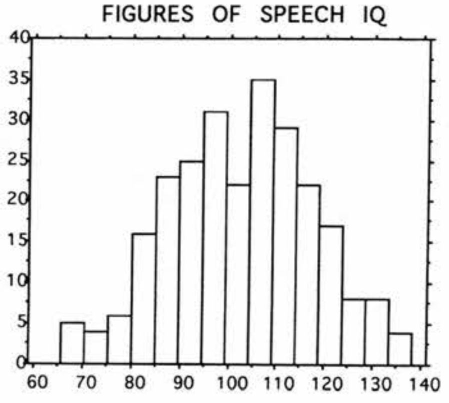
"LOW"





"HIGH"





APPENDIX 6.3

Correlations Between Standardised Test Scores in Samples Below and Above the Mean IQ of each Variable, with the Relevant Selector Excluded from the Matrix.

Correlations in the 'Low' and 'High' IQ samples appear in the lower and upper triangles, respectively.

SAMPLE SELECTOR: RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
Verbal Meaning		**.21**	**.22**	**.21**	**.28**	**.15**	**.25**	**.13**	**.18**
Number Facility	.31		**.22**	**.24**	**.40**	**.46**	**.33**	**.11**	**.25**
Spatial Relations	.31	.28		**.28**	**.28**	**.34**	**.22**	**.05**	**.20**
Figure Grouping	.19	.28	.30		**.33**	**.14**	**.10**	**-.05**	**.12**
Word Grouping	.42	.30	.23	.31		**.33**	**.25**	**.07**	**.16**
Perceptual Speed	.27	.47	.26	.23	.29		**.29**	**.16**	**.20**
Word Fluency	.31	.29	.11	.05	.21	.30		**.04**	**.34**
Digit Span	.17	.20	.07	.08	.12	.14	.09		**.18**
Figures of Speech	.16	.19	.19	.08	.18	.23	.30	.13	

SAMPLE SELECTOR: VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.39**	**.41**	**.38**	**.28**	**.28**	**.09**	**.14**	**.14**
Number Facility	.42		**.28**	**.31**	**.45**	**.45**	**.33**	**.16**	**.27**
Spatial Relations	.41	.33		**.36**	**.18**	**.34**	**.17**	**.03**	**.17**
Figure Grouping	.41	.32	.34		**.31**	**.24**	**.13**	**-.02**	**.15**
Word Grouping	.40	.28	.36	.38		**.28**	**.21**	**.15**	**.16**
Perceptual Speed	.31	.50	.32	.22	.35		**.33**	**.16**	**.19**
Word Fluency	.23	.34	.16	.06	.22	.26		**.10**	**.30**
Digit Span	.21	.14	.08	.05	.04	.13	.04		**.18**
Figures of Speech	.20	.24	.27	.12	.20	.27	.34	.11	

SAMPLE SELECTOR: NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.29**	**.37**	**.33**	**.37**	**.23**	**.04**	**.13**	**.14**
Verbal Meaning	.41		**.28**	**.23**	**.36**	**.12**	**.29**	**.18**	**.24**
Spatial Relations	.43	.35		**.34**	**.32**	**.25**	**.18**	**.04**	**.28**
Figure Grouping	.38	.24	.35		**.33**	**.12**	**.01**	**-.01**	**.13**
Word Grouping	.27	.35	.22	.33		**.26**	**.13**	**.13**	**.14**
Perceptual Speed	.20	.24	.32	.22	.27		**.16**	**.14**	**.17**
Word Fluency	.21	.26	.11	.13	.24	.31		**-.04**	**.36**
Digit Span	.19	.10	.06	.05	.04	.13	.15		**.16**
Figures of Speech	.14	.12	.13	.10	.17	.20	.25	.12	

APPENDIX 6.3 Ctd.

Inter-test correlations in samples above (upper triangle) and below (lower triangle) mean IQ on each variable.

SAMPLE SELECTOR: SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.30**	**.40**	**.40**	**.35**	**.35**	**.21**	**.21**	**.21**
Verbal Meaning	.40		**.26**	**.24**	**.27**	**.22**	**.27**	**.19**	**.21**
Number Facility	.40	.35		**.31**	**.41**	**.46**	**.38**	**.14**	**.37**
Figure Grouping	.36	.24	.34		**.40**	**.23**	**.12**	**.05**	**.15**
Word Grouping	.38	.48	.36	.32		**.35**	**.19**	**.12**	**.14**
Perceptual Speed	.20	.20	.47	.20	.30		**.26**	**.16**	**.26**
Word Fluency	.15	.32	.31	.10	.29	.33		**.11**	**.38**
Digit Span	.17	.12	.20	.04	.12	.17	.06		**.16**
Figures of Speech	.11	.16	.14	.09	.21	.16	.28	.15	

SAMPLE SELECTOR: FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
Raven's Matrices		**.34**	**.40**	**.45**	**.37**	**.30**	**.22**	**.19**	**.22**
Verbal Meaning	.40		**.28**	**.31**	**.34**	**.23**	**.36**	**.21**	**.26**
Number Facility	.41	.39		**.30**	**.37**	**.48**	**.37**	**.20**	**.32**
Spatial Relations	.36	.34	.32		**.35**	**.36**	**.20**	**.20**	**.25**
Word Grouping	.32	.42	.38	.20		**.29**	**.28**	**.09**	**.19**
Perceptual Speed	.30	.25	.48	.31	.34		**.24**	**.19**	**.23**
Word Fluency	.19	.30	.38	.21	.24	.38		**.12**	**.41**
Digit Span	.23	.14	.16	.04	.21	.19	.12		**.25**
Figures of Speech	.13	.13	.21	.18	.16	.23	.25	.11	

SAMPLE SELECTOR: WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
Raven's Matrices		**.25**	**.41**	**.45**	**.39**	**.32**	**.17**	**.12**	**.20**
Verbal Meaning	.40		**.24**	**.16**	**.21**	**.12**	**.24**	**.15**	**.18**
Number Facility	.33	.28		**.26**	**.25**	**.47**	**.32**	**.19**	**.30**
Spatial Relations	.37	.43	.29		**.40**	**.32**	**.13**	**.04**	**.20**
Figure Grouping	.37	.22	.34	.26		**.18**	**.06**	**-.05**	**.16**
Perceptual Speed	.19	.21	.42	.28	.18		**.18**	**.18**	**.18**
Word Fluency	.15	.32	.33	.21	.12	.37		**.09**	**.37**
Digit Span	.20	.08	.11	.09	.10	.12	.08		**.14**
Figures of Speech	.12	.20	.19	.23	.09	.26	.27	.19	

APPENDIX 6.3 Ctd.

Inter-test correlations in samples above (upper triangle) and below (lower triangle) mean IQ on each variable.

SAMPLE SELECTOR: PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
Raven's Matrices		**.35**	**.41**	**.49**	**.40**	**.45**	**.13**	**.19**	**.18**
Verbal Meaning	.42		**.27**	**.31**	**.26**	**.40**	**.32**	**.21**	**.22**
Number Facility	.40	.37		**.23**	**.31**	**.37**	**.23**	**.14**	**.24**
Spatial Relations	.36	.35	.35		**.37**	**.33**	**.11**	**.03**	**.23**
Figure Grouping	.42	.29	.33	.37		**.37**	**.01**	**.04**	**.14**
Word Grouping	.31	.40	.35	.25	.37		**.12**	**.14**	**.09**
Word Fluency	.19	.26	.37	.22	.17	.31		**.06**	**.28**
Digit Span	.18	.11	.20	.13	.06	.11	.08		**.13**
Figures of Speech	.18	.18	.23	.20	.14	.24	.36	.18	

SAMPLE SELECTOR: WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
Raven's Matrices		**.38**	**.43**	**.51**	**.44**	**.41**	**.31**	**.16**	**.23**
Verbal Meaning	.40		**.33**	**.30**	**.35**	**.41**	**.25**	**.19**	**.20**
Number Facility	.43	.30		**.34**	**.36**	**.44**	**.44**	**.10**	**.35**
Spatial Relations	.40	.36	.32		**.39**	**.30**	**.32**	**.08**	**.27**
Figure Grouping	.42	.21	.37	.36		**.38**	**.20**	**.04**	**.21**
Word Grouping	.39	.36	.35	.34	.40		**.25**	**.16**	**.25**
Perceptual Speed	.32	.19	.49	.35	.33	.43		**.14**	**.25**
Digit Span	.25	.12	.27	.10	.07	.09	.21		**.19**
Figures of Speech	.11	.15	.09	.16	.08	.08	.13	.11	

SAMPLE SELECTOR: DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
Raven's Matrices		**.30**	**.38**	**.44**	**.35**	**.36**	**.31**	**.12**	**.16**
Verbal Meaning	.48		**.30**	**.31**	**.25**	**.42**	**.27**	**.35**	**.24**
Number Facility	.49	.38		**.30**	**.30**	**.40**	**.45**	**.31**	**.27**
Spatial Relations	.48	.44	.40		**.41**	**.33**	**.34**	**.20**	**.24**
Figure Grouping	.46	.30	.37	.35		**.29**	**.29**	**.12**	**.16**
Word Grouping	.46	.41	.42	.35	.47		**.38**	**.26**	**.19**
Perceptual Speed	.37	.27	.51	.41	.23	.36		**.31**	**.23**
Word Fluency	.30	.31	.43	.26	.19	.28	.33		**.39**
Figures of Speech	.19	.18	.27	.26	.14	.19	.21	.28	

APPENDIX 6.3 Ctd.

Inter-test correlations in samples above (upper triangle) and below (lower triangle) mean IQ on each variable.

SAMPLE SELECTOR: FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
Raven's Matrices		**.39**	**.44**	**.49**	**.46**	**.40**	**.34**	**.16**	**.16**
Verbal Meaning	.41		**.37**	**.34**	**.28**	**.40**	**.27**	**.27**	**.18**
Number Facility	.44	.33		**.39**	**.34**	**.44**	**.47**	**.39**	**.15**
Spatial Relations	.42	.35	.29		**.38**	**.31**	**.37**	**.21**	**.05**
Figure Grouping	.39	.28	.37	.38		**.39**	**.21**	**.10**	**.03**
Word Grouping	.42	.41	.38	.33	.38		**.25**	**.26**	**.09**
Perceptual Speed	.31	.25	.50	.32	.30	.45		**.25**	**.13**
Word Fluency	.21	.34	.29	.16	.14	.21	.31		**.08**
Digit Span	.25	.14	.18	.10	.05	.13	.19	.03	

APPENDIX 6.4

TEST RELIABILITIES IN 'LOW' AND 'HIGH' IQ SAMPLES SELECTED USING EACH TEST AS THE CRITERION.

TEST RELIABILITIES IN TEN 'LOW' IQ SAMPLES

SELECTOR:	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.92	.92	.93	.93	.93	.94	.93	.86	.92
Verbal Meaning	.94	*	.95	.95	.94	.95	.95	.95	.90	.95
Number Facility	.95	.96	*	.96	.95	.96	.95	.96	.92	.96
Spatial Relations	.81	.80	.84	*	.82	.82	.83	.83	.70	.86
Figure Grouping	.78	.76	.78	.75	*	.73	.77	.72	.63	.79
Word Grouping	.81	.82	.80	.83	.80	*	.79	.82	.64	.80
Perceptual Speed	.95	.95	.95	.95	.95	.95	*	.95	.91	.95
Word Fluency	.96	.96	.96	.96	.96	.96	.95	*	.93	.96
Digit Span	.92	.92	.93	.92	.93	.92	.92	.92	*	.92
Figures of Speech	.79	.79	.80	.80	.78	.80	.80	.80	.68	*

TEST RELIABILITIES IN TEN 'HIGH' IQ SAMPLES

SELECTOR:	RPM	VM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices	*	.91	.90	.89	.90	.90	.90	.91	.85	.92
Verbal Meaning	.95	*	.94	.94	.95	.94	.94	.94	.91	.95
Number Facility	.96	.95	*	.95	.96	.95	.96	.95	.93	.95
Spatial Relations	.82	.83	.78	*	.81	.81	.80	.82	.71	.78
Figure Grouping	.63	.71	.61	.71	*	.70	.68	.76	.52	.69
Word Grouping	.69	.68	.68	.68	.73	*	.73	.71	.64	.77
Perceptual Speed	.96	.95	.95	.95	.96	.95	*	.96	.93	.96
Word Fluency	.96	.96	.96	.96	.97	.96	.96	*	.92	.96
Digit Span	.92	.92	.92	.92	.92	.92	.92	.92	*	.92
Figures of Speech	.82	.82	.82	.82	.84	.82	.82	.80	.71	*

APPENDIX 6.5

Inter-test correlations in samples above **(upper triangle)** and below (lower triangle) mean IQ on each variable, corrected for test unreliability.

SAMPLE SELECTOR: RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
Verbal Meaning		**.22**	**.25**	**.27**	**.35**	**.15**	**.27**	**.14**	**.20**
Number Facility	.32		**.25**	**.31**	**.49**	**.48**	**.35**	**.12**	**.28**
Spatial Relations	.36	.32		**.40**	**.37**	**.38**	**.24**	**.05**	**.25**
Figure Grouping	.22	.33	.38		**.50**	**.18**	**.13**	**-.06**	**.16**
Word Grouping	.48	.34	.28	.39		**.41**	**.30**	**.09**	**.21**
Perceptual Speed	.28	.50	.30	.27	.33		**.30**	**.17**	**.22**
Word Fluency	.32	.30	.12	.05	.24	.32		**.04**	**.38**
Digit Span	.18	.21	.08	.09	.14	.15	.10		**.20**
Figures of Speech	.19	.22	.23	.10	.22	.26	.35	.16	

SAMPLE SELECTOR: VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.42**	**.47**	**.48**	**.36**	**.30**	**.09**	**.16**	**.17**
Number Facility	.45		**.32**	**.37**	**.55**	**.47**	**.35**	**.18**	**.30**
Spatial Relations	.48	.37		**.46**	**.24**	**.38**	**.19**	**.03**	**.21**
Figure Grouping	.49	.38	.44		**.45**	**.29**	**.16**	**-.02**	**.19**
Word Grouping	.46	.31	.44	.48		**.35**	**.26**	**.18**	**.21**
Perceptual Speed	.33	.53	.36	.26	.40		**.35**	**.17**	**.22**
Word Fluency	.24	.35	.19	.07	.24	.28		**.11**	**.34**
Digit Span	.22	.15	.09	.06	.05	.14	.04		**.21**
Figures of Speech	.24	.28	.34	.16	.25	.32	.39	.13	

SAMPLE SELECTOR: NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.31**	**.45**	**.45**	**.47**	**.25**	**.05**	**.14**	**.17**
Verbal Meaning	.43		**.33**	**.30**	**.46**	**.13**	**.31**	**.20**	**.27**
Spatial Relations	.49	.39		**.49**	**.44**	**.29**	**.21**	**.05**	**.35**
Figure Grouping	.45	.28	.43		**.52**	**.16**	**.02**	**-.01**	**.18**
Word Grouping	.31	.40	.27	.41		**.32**	**.16**	**.17**	**.19**
Perceptual Speed	.22	.25	.36	.25	.31		**.16**	**.15**	**.20**
Word Fluency	.22	.27	.12	.15	.28	.33		**-.04**	**.41**
Digit Span	.21	.10	.06	.06	.05	.13	.15		**.18**
Figures of Speech	.16	.14	.16	.12	.21	.23	.29	.13	

APPENDIX 6.5 ctd.

Inter-test correlations in samples above (**upper triangle**) and below (lower triangle) mean IQ on each variable, corrected for test unreliability.

SAMPLE SELECTOR: SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
Raven's Matrices		**.33**	**.44**	**.50**	**.45**	**.38**	**.22**	**.24**	**.25**
Verbal Meaning	.42		**.27**	**.30**	**.34**	**.24**	**.28**	**.21**	**.24**
Number Facility	.42	.37		**.37**	**.50**	**.49**	**.40**	**.15**	**.42**
Figure Grouping	.43	.28	.40		**.57**	**.27**	**.15**	**.07**	**.20**
Word Grouping	.43	.54	.40	.40		**.43**	**.24**	**.15**	**.19**
Perceptual Speed	.22	.21	.49	.24	.33		**.27**	**.17**	**.30**
Word Fluency	.16	.34	.32	.11	.33	.34		**.12**	**.42**
Digit Span	.18	.12	.22	.04	.14	.18	.06		**.18**
Figures of Speech	.13	.19	.16	.12	.26	.18	.32	.17	

SAMPLE SELECTOR: FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
Raven's Matrices		**.37**	**.43**	**.53**	**.46**	**.32**	**.24**	**.21**	**.26**
Verbal Meaning	.43		**.29**	**.35**	**.40**	**.24**	**.37**	**.23**	**.29**
Number Facility	.43	.41		**.34**	**.44**	**.50**	**.38**	**.22**	**.35**
Spatial Relations	.42	.39	.36		**.46**	**.41**	**.22**	**.23**	**.30**
Word Grouping	.38	.49	.44	.25		**.35**	**.34**	**.10**	**.25**
Perceptual Speed	.32	.27	.50	.35	.39		**.25**	**.20**	**.26**
Word Fluency	.20	.31	.40	.23	.27	.40		**.12**	**.45**
Digit Span	.25	.15	.17	.04	.24	.20	.13		**.29**
Figures of Speech	.15	.15	.24	.23	.20	.26	.29	.13	

SAMPLE SELECTOR: WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
Raven's Matrices		**.27**	**.44**	**.52**	**.49**	**.35**	**.18**	**.13**	**.23**
Verbal Meaning	.43		**.25**	**.19**	**.26**	**.13**	**.25**	**.16**	**.21**
Number Facility	.35	.30		**.29**	**.31**	**.49**	**.33**	**.20**	**.34**
Spatial Relations	.43	.49	.33		**.53**	**.37**	**.15**	**.05**	**.25**
Figure Grouping	.45	.27	.41	.34		**.22**	**.07**	**-.07**	**.21**
Perceptual Speed	.20	.22	.43	.31	.22		**.19**	**.19**	**.20**
Word Fluency	.15	.34	.34	.24	.15	.38		**.09**	**.42**
Digit Span	.21	.09	.12	.10	.12	.12	.09		**.16**
Figures of Speech	.14	.23	.22	.28	.12	.30	.31	.22	

APPENDIX 6.5 Ctd.

Inter-test correlations in samples above (**upper triangle**) and below (lower triangle) mean IQ on each variable, corrected for test unreliability.

SAMPLE SELECTOR: PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
Raven's Matrices		**.38**	**.44**	**.58**	**.51**	**.55**	**.14**	**.21**	**.21**
Verbal Meaning	.45		**.28**	**.35**	**.32**	**.48**	**.34**	**.23**	**.25**
Number Facility	.43	.39		**.26**	**.39**	**.44**	**.24**	**.15**	**.27**
Spatial Relations	.41	.40	.40		**.49**	**.43**	**.12**	**.03**	**.29**
Figure Grouping	.49	.34	.38	.46		**.52**	**.01**	**.04**	**.19**
Word Grouping	.36	.46	.40	.31	.48		**.14**	**.17**	**.12**
Word Fluency	.20	.28	.39	.25	.20	.35		**.07**	**.31**
Digit Span	.20	.11	.21	.14	.07	.12	.09		**.15**
Figures of Speech	.21	.21	.26	.24	.17	.30	.41	.21	

SAMPLE SELECTOR: WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
Raven's Matrices		**.41**	**.47**	**.59**	**.53**	**.51**	**.34**	**.17**	**.27**
Verbal Meaning	.42		**.35**	**.35**	**.42**	**.51**	**.26**	**.21**	**.23**
Number Facility	.46	.31		**.38**	**.42**	**.53**	**.46**	**.11**	**.40**
Spatial Relations	.45	.41	.36		**.50**	**.39**	**.36**	**.09**	**.34**
Figure Grouping	.51	.26	.45	.47		**.52**	**.24**	**.05**	**.27**
Word Grouping	.44	.40	.39	.41	.52		**.30**	**.20**	**.34**
Perceptual Speed	.34	.20	.52	.40	.40	.48		**.15**	**.29**
Digit Span	.27	.12	.29	.12	.08	.10	.22		**.22**
Figures of Speech	.13	.17	.10	.20	.10	.10	.15	.13	

SAMPLE SELECTOR: DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
Raven's Matrices		**.34**	**.43**	**.57**	**.52**	**.49**	**.35**	**.13**	**.20**
Verbal Meaning	.55		**.33**	**.40**	**.37**	**.56**	**.29**	**.38**	**.30**
Number Facility	.55	.42		**.38**	**.43**	**.52**	**.49**	**.34**	**.33**
Spatial Relations	.62	.55	.50		**.67**	**.48**	**.41**	**.24**	**.34**
Figure Grouping	.63	.40	.49	.53		**.50**	**.41**	**.17**	**.27**
Word Grouping	.62	.55	.55	.52	.74		**.49**	**.34**	**.29**
Perceptual Speed	.41	.30	.55	.51	.31	.47		**.33**	**.28**
Word Fluency	.34	.34	.47	.33	.24	.37	.35		**.48**
Figures of Speech	.25	.23	.34	.37	.21	.28	.27	.35	

APPENDIX 6.5 Ctd.

Inter-test correlations in samples above (**upper triangle**) and below (lower triangle) mean IQ on each variable, corrected for test unreliability.

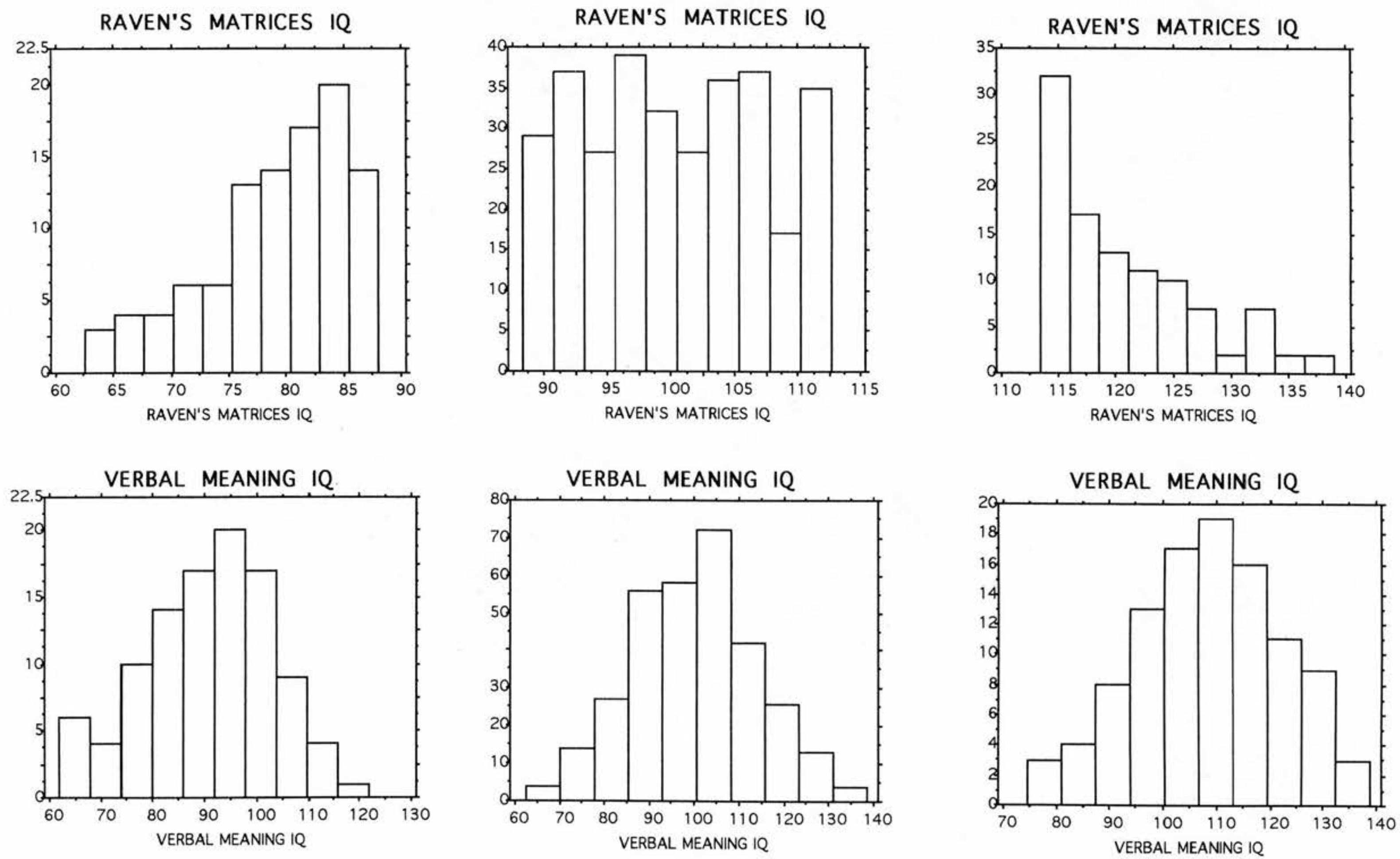
SAMPLE SELECTOR: FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
Raven's Matrices		**.42**	**.47**	**.58**	**.58**	**.48**	**.36**	**.17**	**.17**
Verbal Meaning	.44		**.39**	**.40**	**.35**	**.47**	**.28**	**.28**	**.19**
Number Facility	.46	.35		**.45**	**.42**	**.51**	**.49**	**.41**	**.16**
Spatial Relations	.47	.38	.32		**.51**	**.41**	**.43**	**.24**	**.06**
Figure Grouping	.46	.33	.42	.46		**.54**	**.26**	**.12**	**.04**
Word Grouping	.49	.47	.43	.40	.48		**.29**	**.30**	**.10**
Perceptual Speed	.33	.26	.52	.36	.35	.52		**.26**	**.14**
Word Fluency	.22	.35	.30	.17	.16	.24	.32		**.09**
Digit Span	.27	.15	.19	.11	.06	.15	.21	.04	

APPENDIX 6.6 DISTRIBUTIONS OF TEST VARIABLES IN SUB-GROUPS SELECTED BY RAVEN'S MATRICES IQ.

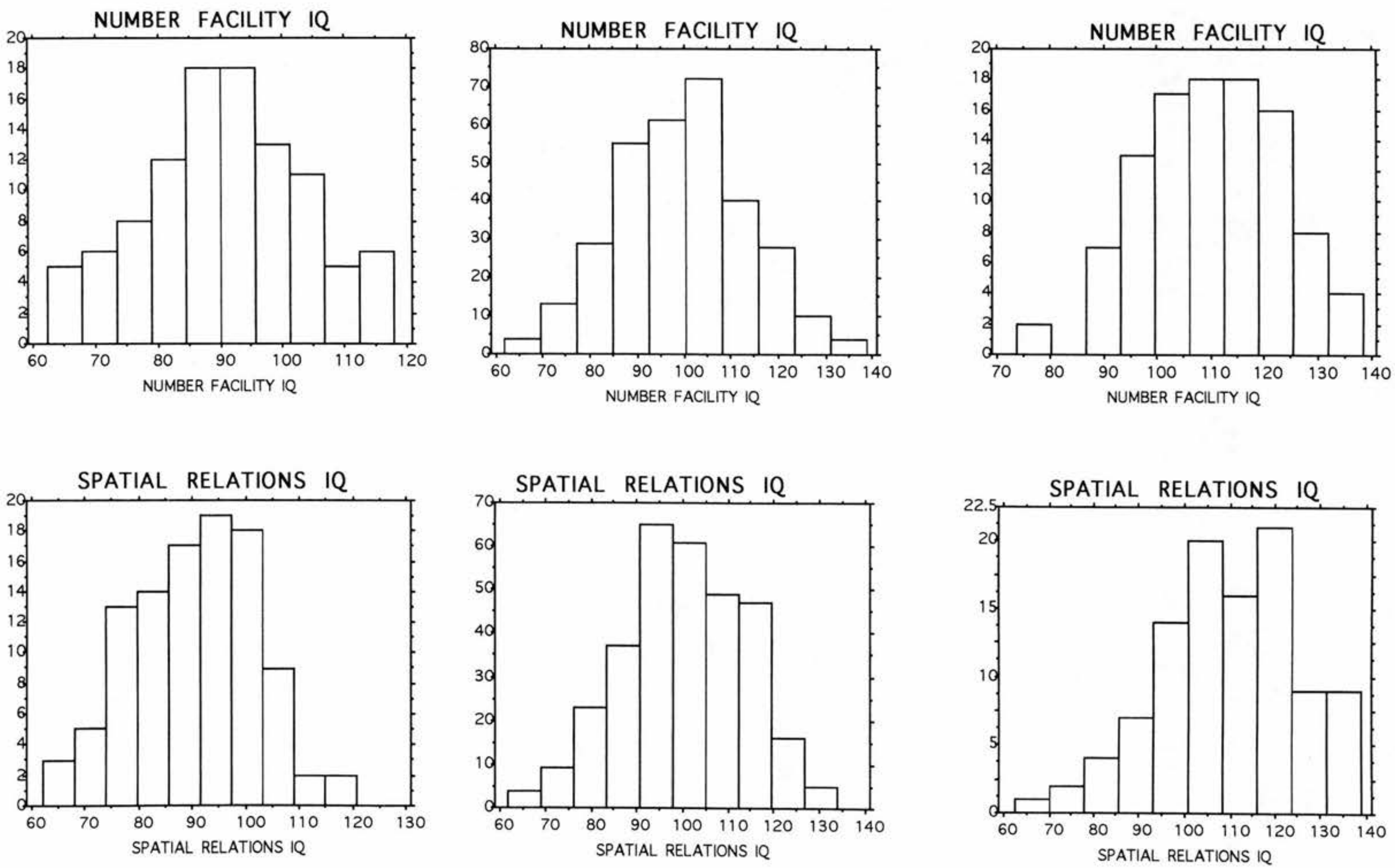
['LOW': RPM IQ RANGE = 62-87, MEAN=94; 'MID' RPM RANGE =IQ 87.5-112.5, MEAN=100, 'HIGH' RPM IQ RANGE= 113-138, MEAN=106.]

"LOW" "MID" "HIGH"



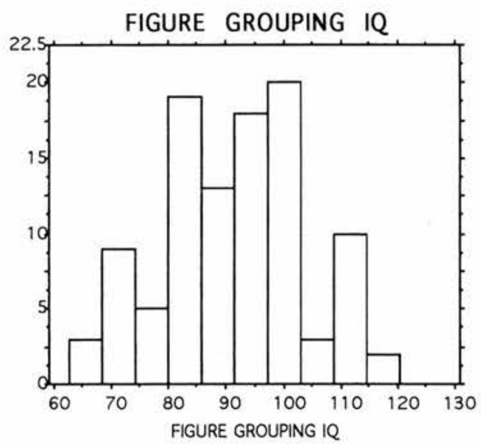
APPENDIX 6.6 CTD. DISTRIBUTIONS OF TEST VARIABLES IN SUB-GROUPS SELECTED BY RAVEN'S MATRICES IQ.

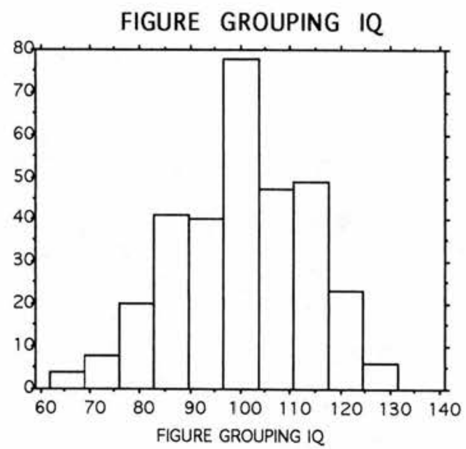
"LOW" "MID" "HIGH"

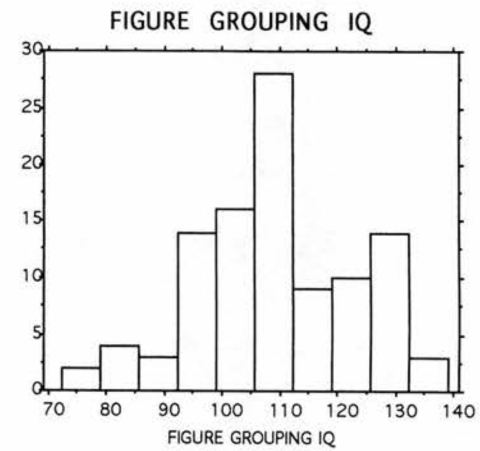


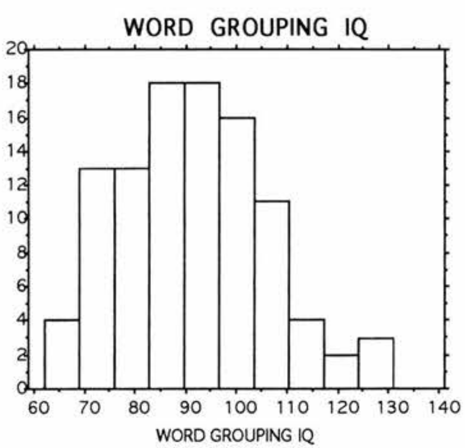
APPENDIX 6.6 CTD. DISTRIBUTIONS OF TEST VARIABLES IN SUB-GROUPS SELECTED BY RAVEN'S MATRICES IQ.

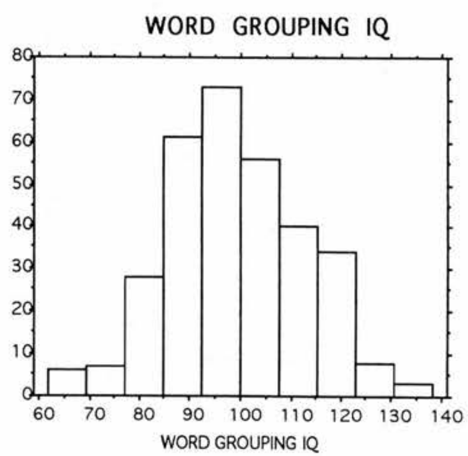
"LOW" "MID" "HIGH"

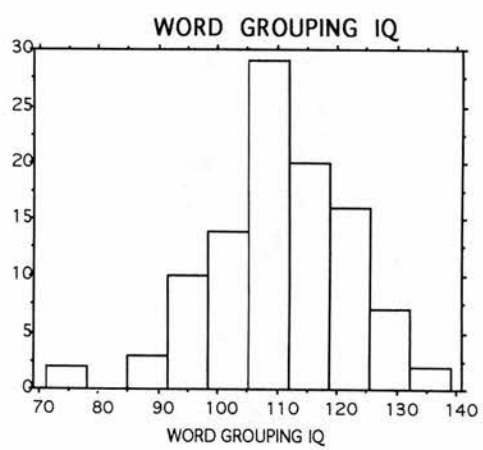




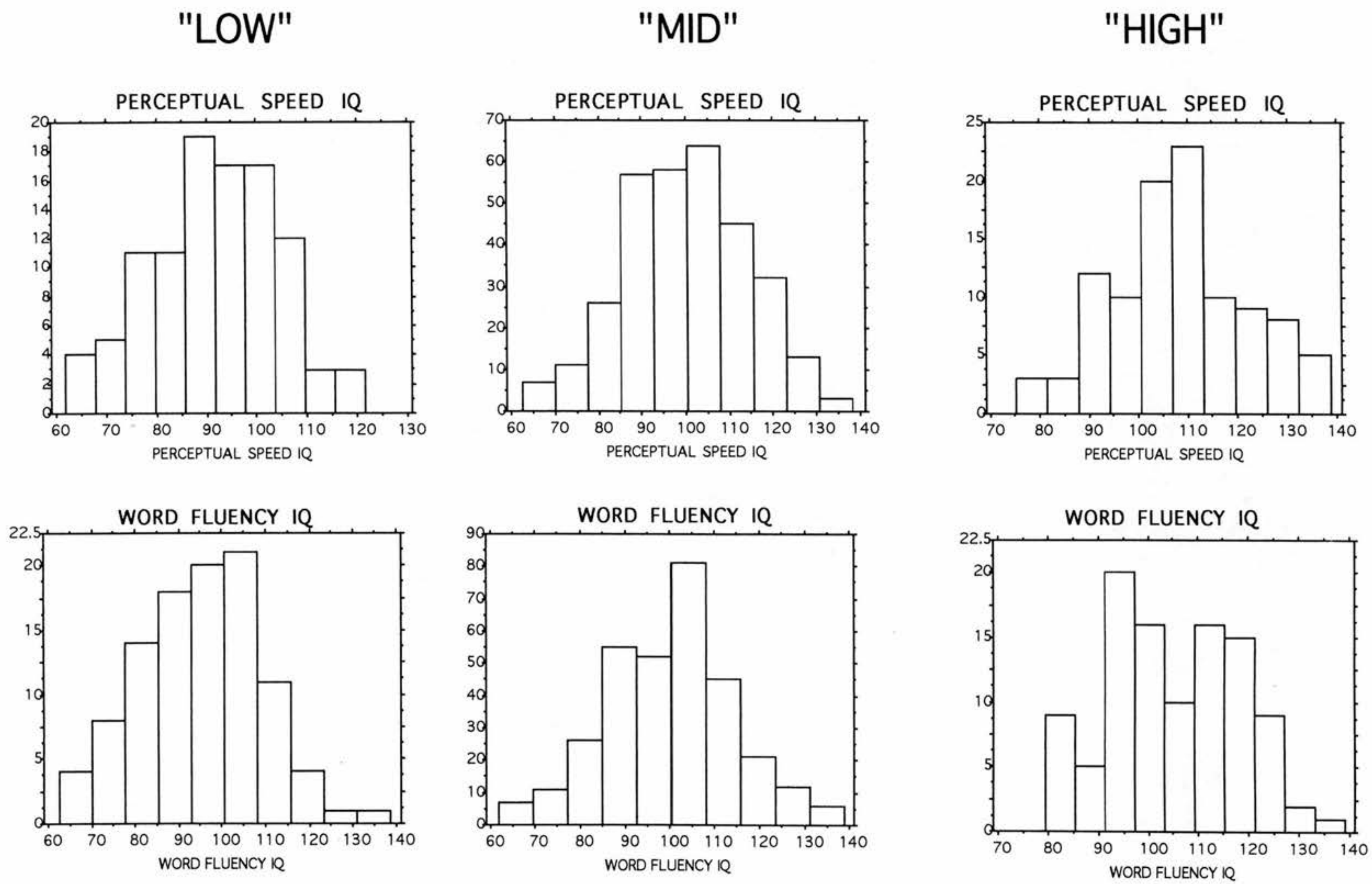




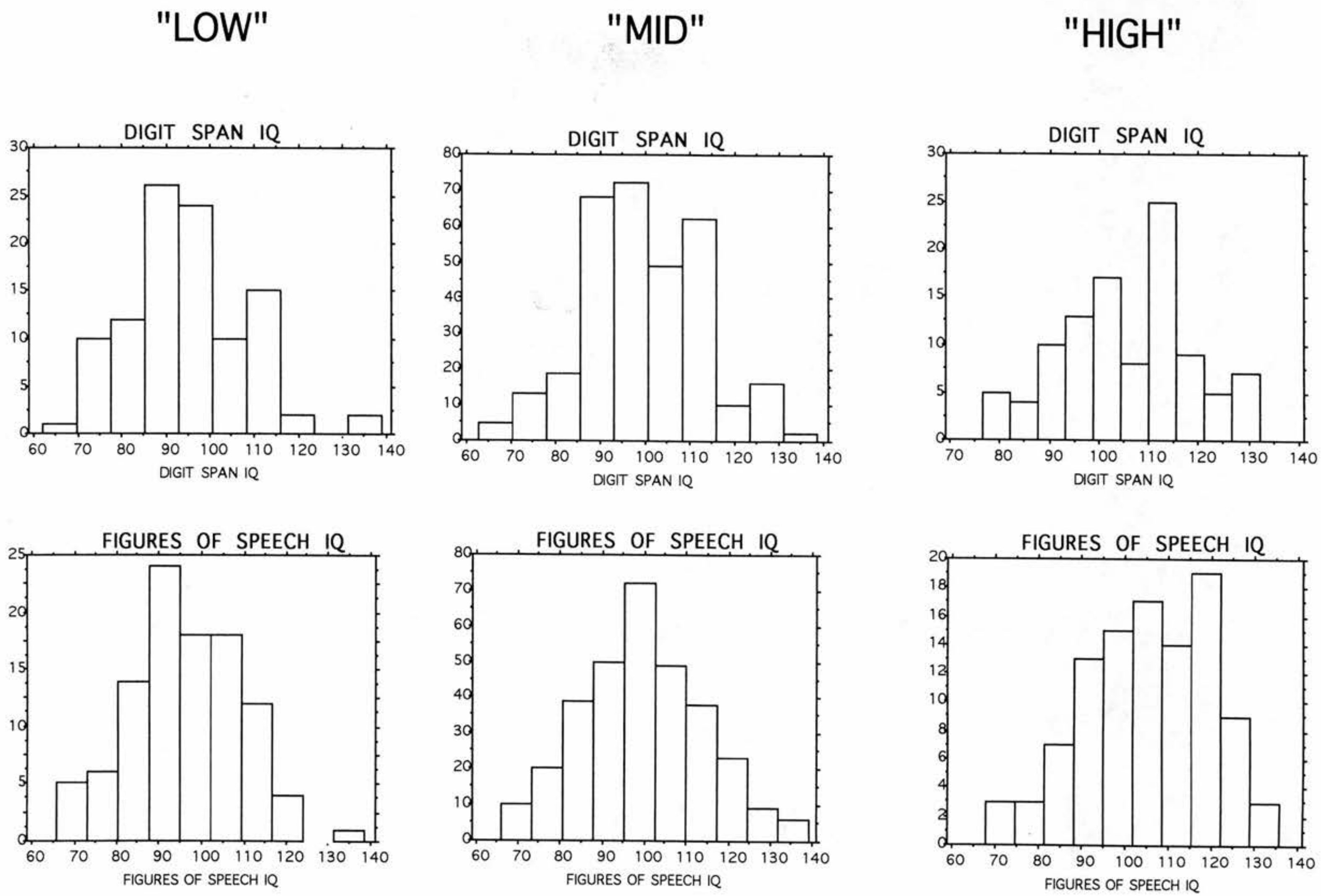




APPENDIX 6.6 CTD. DISTRIBUTIONS OF TEST VARIABLES IN SUB-GROUPS SELECTED BY RAVEN'S MATRICES IQ.



APPENDIX 6.6 CTD. DISTRIBUTIONS OF TEST VARIABLES IN SUB-GROUPS SELECTED BY RAVEN'S MATRICES IQ.



APPENDIX 6.7

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, SELECTED ACCORDING TO IQ ON EACH TEST, WITH THE RELEVANT SELECTOR EXCLUDED FROM THE MATRIX.

'LOW' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	1.0	.33	.34	.18	.35	.22	.42	.11	.06
NF	.33	1.0	.21	.17	.15	.40	.42	.24	.25
SR	.34	.21	1.0	.21	.12	.09	.13	.03	.11
FG	.18	.17	.21	1.0	.32	.09	.12	.18	.11
WG	.35	.15	.12	.32	1.0	.21	.28	.12	.21
PS	.22	.40	.09	.09	.21	1.0	.41	.11	.18
WF	.42	.42	.13	.12	.28	.41	1.0	.12	.18
DS	.11	.24	.03	.18	.12	.11	.12	1.0	.05
FS	.06	.25	.11	.11	.21	.18	.18	.05	1.0

'MID' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	1.0	.24	.25	.15	.34	.19	.23	.08	.22
NF	.24	1.0	.27	.32	.37	.42	.32	.04	.23
SR	.25	.27	1.0	.23	.25	.35	.16	.00	.24
FG	.15	.32	.23	1.0	.30	.22	.06	-.06	.07
WG	.34	.37	.25	.30	1.0	.29	.19	.01	.15
PS	.19	.42	.35	.22	.29	1.0	.23	.11	.22
WF	.23	.32	.16	.06	.19	.23	1.0	.07	.36
DS	.08	.04	.00	-.06	.01	.11	.07	1.0	.11
FS	.22	.23	.24	.07	.15	.22	.36	.11	1.0

'HIGH' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	.01	.16	.19	.26	.21	.08	.31	.05	.12
NF	.16	1.0	.18	.07	.27	.5	.28	.13	.23
SR	.19	.18	1.0	.42	.26	.21	.18	.00	.13
FG	.26	.07	.42	1.0	.26	.02	.07	-.16	.15
WG	.21	.27	.26	.26	1.0	.35	.25	.08	.11
PS	.08	.50	.21	.02	.35	1.0	.32	.13	.18
WF	.31	.28	.18	.07	.25	.32	1.0	-.06	.33
DS	.05	.13	.00	-.16	.08	.13	-.06	1.0	.24
FS	.12	.23	.13	.15	.11	.18	.33	.24	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.39	.47	.39	.47	.34	.21	.19	.13
NF	.39	1.0	.34	.39	.35	.59	.29	.20	.21
SR	.47	.34	1.0	.36	.44	.35	.13	.02	.13
FG	.39	.39	.36	1.0	.45	.39	.07	-.01	.14
WG	.47	.35	.44	.45	1.0	.36	.25	.02	.19
PS	.34	.59	.35	.39	.36	1.0	.18	.12	.17
WF	.21	.29	.13	.07	.25	.18	1.0	-.02	.42
DS	.19	.20	.02	-.01	.02	.12	-.02	1.0	.16
FS	.13	.21	.13	.14	.19	.17	.42	.16	1.0

'MID' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.40	.38	.35	.32	.28	.09	.20	.11
NF	.40	1.0	.29	.33	.35	.46	.29	.14	.17
SR	.38	.29	1.0	.31	.24	.33	.15	.10	.20
FG	.35	.33	.31	1.0	.33	.18	.04	.04	.06
WG	.32	.35	.24	.33	1.0	.35	.19	.12	.13
PS	.28	.46	.33	.18	.35	1.0	.30	.16	.20
WF	.09	.29	.15	.04	.19	.30	1.0	.10	.25
DS	.20	.14	.10	.04	.12	.16	.10	1.0	.09
FS	.11	.17	.20	.06	.13	.20	.25	.09	1.0

'HIGH' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.37	.38	.44	.22	.25	.20	.01	.20
NF	.37	1.0	.23	.20	.41	.37	.39	.15	.40
SR	.38	.23	1.0	.38	.14	.25	.12	-.07	.19
FG	.44	.20	.38	1.0	.22	.19	.18	.01	.21
WG	.22	.41	.14	.22	1.0	.17	.14	.09	.15
PS	.25	.37	.25	.19	.17	1.0	.27	.19	.24
WF	.20	.39	.12	.18	.14	.27	1.0	.03	.33
DS	.01	.15	-.07	.01	.09	.19	.03	1.0	.23
FS	.20	.40	.19	.21	.15	.24	.33	.23	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.36	.15	.17	.19	.06	.05	.21	.16
VM	.36	1.0	.26	.22	.31	.22	.34	.18	.10
SR	.15	.26	1.0	.38	.20	.29	.10	.06	.06
FG	.17	.22	.38	1.0	.43	.12	.12	-.04	.16
WG	.19	.31	.20	.43	1.0	.28	.20	.00	.07
PS	.06	.22	.29	.12	.28	1.0	.34	.15	.14
WF	.05	.34	.10	.12	.20	.34	1.0	.20	.18
DS	.21	.18	.06	-.04	.00	.15	.20	1.0	.12
FS	.16	.10	.06	.16	.07	.14	.18	.12	1.0

'MID' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.35	.46	.45	.39	.22	.12	.18	.10
VM	.35	1.0	.35	.25	.39	.21	.19	.12	.15
SR	.46	.35	1.0	.31	.29	.27	.12	.10	.17
FG	.45	.25	.31	1.0	.29	.17	.03	.06	.06
WG	.39	.39	.29	.29	1.0	.25	.15	.19	.16
PS	.22	.21	.27	.17	.25	1.0	.25	.14	.20
WF	.12	.19	.12	.03	.15	.25	1.0	.09	.28
DS	.18	.12	.10	.06	.19	.14	.09	1.0	.12
FS	.10	.15	.17	.06	.16	.20	.28	.12	1.0

'HIGH' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	1.0	.28	.41	.22	.24	.33	.14	.06	.25
VM	.28	1.0	.25	.22	.33	.07	.45	.13	.41
SR	.41	.25	1.0	.34	.33	.37	.25	-.09	.36
FG	.22	.22	.34	1.0	.32	.17	.07	-.1	.14
WG	.24	.33	.33	.32	1.0	.34	.3	-.11	.23
PS	.33	.07	.37	.17	.34	1.0	.13	.08	.15
WF	.14	.45	.25	.07	.3	.13	1.0	-.08	.47
DS	.06	.13	-.09	-.1	-.11	.08	-.08	1.0	.14
FS	.25	.41	.36	.14	.23	.15	.47	.14	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	1.0	.36	.28	.31	.25	.09	.09	.28	.11
VM	.36	1.0	.32	.15	.51	.12	.3	.19	.11
NF	.28	.32	1.0	.34	.23	.45	.25	.24	.16
FG	.31	.15	.34	1.0	.25	.22	.09	.03	.12
WG	.25	.51	.23	.25	10	.27	.3	.17	.27
PS	.09	.12	.45	.22	.27	1.0	.25	.20	.12
WF	.09	.30	.25	.09	.30	.25	1.0	.16	.21
DS	.28	.19	.24	.03	.17	.20	.16	1.0	.23
FS	.11	.11	.16	.12	.27	.12	.21	.23	1.0

'MID' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	1.0	.35	.45	.32	.37	.31	.16	.14	.15
VM	.35	1.0	.32	.23	.35	.21	.30	.15	.21
NF	.45	.32	1.0	.25	.41	.49	.34	.17	.20
FG	.32	.23	.25	1.0	.35	.14	.01	.04	.02
WG	.37	.35	.41	.35	1.0	.33	.24	.10	.12
PS	.31	.21	.49	.14	.33	1.0	.32	.17	.22
WF	.16	.30	.34	.01	.24	.32	1.0	.10	.34
DS	.14	.15	.17	.04	.10	.17	.10	1.0	.11
FS	.15	.21	.2	.02	.12	.22	.34	.11	1.0

'HIGH' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	1.0	.25	.26	.50	.41	.31	.16	.21	.04
VM	.25	1.0	.18	.28	.25	.25	.20	.09	.11
NF	.26	.18	1.0	.40	.39	.38	.40	.08	.38
FG	.50	.28	.40	1.0	.37	.27	.22	.04	.26
WG	.41	.25	.39	.37	1.0	.33	.1.0	.11	.15
PS	.31	.25	.38	.27	.33	1.0	.18	.13	.19
WF	.16	.20	.40	.22	.10	.18	1.0	-.05	.32
DS	.21	.09	.08	.04	.11	.13	-.05	1.0	.17
FS	.04	.11	.38	.26	.15	.19	.32	.17	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	1.0	.24	.33	.29	.21	.22	.25	.19	.05
VM	.24	1.0	.35	.26	.40	.29	.30	.16	.07
NF	.33	.35	1.0	.40	.39	.44	.51	-.05	.24
SR	.29	.26	.40	1.0	.18	.36	.26	.1.0	.25
WG	.21	.40	.39	.18	1.0	.29	.34	.14	.16
PS	.22	.29	.44	.36	.29	1.0	.53	.04	.25
WF	.25	.30	.51	.26	.34	.53	1.0	.16	.23
DS	.19	.16	-.05	.10	.14	.04	.16	1.0	.01
FS	.05	.07	.24	.25	.16	.25	.23	.01	1.0

'MID' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	1.0	.39	.37	.39	.36	.33	.16	.19	.16
VM	.39	1.0	.25	.30	.36	.21	.30	.14	.21
NF	.37	.25	1.0	.21	.33	.47	.34	.25	.24
SR	.39	.30	.21	1.0	.27	.31	.15	.04	.18
WG	.36	.36	.33	.27	1.0	.33	.25	.12	.19
PS	.33	.21	.47	.31	.33	1.0	.24	.21	.22
WF	.16	.30	.34	.15	.25	.24	1.0	.06	.35
DS	.19	.14	.25	.04	.12	.21	.06	1.0	.18
FS	.16	.21	.24	.18	.19	.22	.35	.18	1.0

'HIGH' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	1.0	.30	.39	.50	.36	.15	.23	.29	.33
VM	.30	1.0	.38	.31	.34	.22	.36	.33	.29
NF	.39	.38	1.0	.34	.36	.47	.31	.19	.33
SR	.50	.31	.34	1.0	.36	.34	.32	.28	.30
WG	.36	.34	.36	.36	1.0	.27	.17	.12	.16
PS	.15	.22	.47	.34	.27	1.0	.32	.26	.29
WF	.23	.36	.31	.32	.17	.32	1.0	.17	.39
DS	.29	.33	.19	.28	.12	.26	.17	1.0	.19
FS	.33	.29	.33	.30	.16	.29	.39	.19	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY IQ GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	1.0	.42	.31	.40	.36	.20	.15	.22	.15
VM	.42	1.0	.23	.48	.19	.15	.36	.05	.20
NF	.31	.23	1.0	.27	.47	.43	.21	.26	.20
SR	.40	.48	.27	1.0	.29	.27	.26	.05	.31
FG	.36	.19	.47	.29	1.0	.34	.12	.18	.17
PS	.20	.15	.43	.27	.34	1.0	.28	.16	.20
WF	.15	.36	.21	.26	.12	.28	1.0	.14	.23
DS	.22	.05	.26	.05	.18	.16	.14	1.0	.08
FS	.15	.20	.20	.31	.17	.20	.23	.08	1.0

'MID' ABILITY GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	1.0	.33	.43	.36	.34	.26	.15	.19	.15
VM	.33	1.0	.32	.28	.23	.24	.24	.15	.13
NF	.43	.32	1.0	.31	.23	.47	.36	.16	.23
SR	.36	.28	.31	1.0	.32	.35	.16	.09	.20
FG	.34	.23	.23	.32	1.0	.14	.04	.05	.04
PS	.26	.24	.47	.35	.14	1.0	.32	.15	.25
WF	.15	.24	.36	.16	.04	.32	1.0	.05	.32
DS	.19	.15	.16	.09	.05	.15	.05	1.0	.20
FS	.15	.13	.23	.20	.04	.25	.32	.20	1.0

'HIGH' ABILITY GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	1.0	.18	.28	.54	.37	.37	.10	.11	.12
VM	.18	1.0	.18	.13	.10	.03	.28	.19	.30
NF	.28	.18	1.0	.26	.27	.44	.33	.07	.30
SR	.54	.13	.26	1.0	.39	.23	.12	.00	.17
FG	.37	.10	.27	.39	1.0	.21	.11	-.24	.31
PS	.37	.03	.44	.23	.21	1.0	.07	.21	.04
WF	.1 0	.28	.33	.12	.11	.07	1.0	.12	.37
DS	.11	.19	.07	.00	-.24	.21	.12	1.0	.08
FS	.12	.30	.30	.17	.31	.04	.37	.08	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' IQ GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	1.0	.44	.30	.14	.27	.23	.23	.21	.19
VM	.44	1.0	.44	.29	.30	.35	.27	.13	.26
NF	.30	.44	1.0	.21	.30	.24	.41	.21	.27
SR	.14	.29	.21	1.0	.27	.09	.20	.07	.18
FG	.27	.30	.30	.27	1.0	.37	.26	.05	.13
WG	.23	.35	.24	.09	.37	1.0	.27	.02	.33
WF	.23	.27	.41	.20	.26	.27	1.0	.07	.30
DS	.21	.13	.21	.07	.05	.02	.07	1.0	.28
FS	.19	.26	.27	.18	.13	.33	.30	.28	1.0

'MID' ABILITY GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	1.0	.36	.4	.46	.46	.38	.14	.16	.13
VM	.36	1.0	.29	.30	.23	.40	.27	.11	.14
NF	.40	.29	1.0	.27	.29	.35	.29	.10	.20
SR	.46	.30	.27	1.0	.32	.31	.18	.04	.20
FG	.46	.23	.29	.32	1.0	.36	.06	.01	.11
WG	.38	.40	.35	.31	.36	1.0	.24	.09	.14
WF	.14	.27	.29	.18	.06	.24	1.0	.06	.30
DS	.16	.11	.10	.04	.01	.09	.06	1.0	.07
FS	.13	.14	.20	.20	.11	.14	.30	.07	1.0

'HIGH' ABILITY GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	1.0	.29	.32	.47	.28	.37	.06	.10	.14
VM	.29	1.0	.11	.35	.28	.33	.34	.20	.26
NF	.32	.11	1.0	.19	.32	.35	.13	.08	.17
SR	.47	.35	.19	1.0	.40	.29	-.02	.00	.18
FG	.28	.28	.32	.40	1.0	.28	-.03	.00	.15
WG	.37	.33	.35	.29	.28	1.0	-.02	.14	.03
WF	.06	.34	.13	-.02	-.03	-.02	1.0	-.02	.3
DS	.1	.2	.08	.00	.00	.14	-.02	1.0	.16
FS	.14	.26	.17	.18	.15	.03	.3	.16	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	1.0	.49	.47	.36	.43	.41	.37	.24	.08
VM	.49	1.0	.28	.27	.34	.42	.17	.10	.17
NF	.47	.28	1.0	.26	.39	.29	.63	.37	.11
SR	.36	.27	.26	1.0	.39	.29	.36	.12	.14
FG	.43	.34	.39	.39	1.0	.43	.47	.12	.07
WG	.41	.42	.29	.29	.43	1.0	.46	.10	.01
PS	.37	.17	.63	.36	.47	.46	1.0	.22	.06
DS	.24	.10	.37	.12	.12	.10	.22	1.0	.02
FS	.08	.17	.11	.14	.07	.01	.06	.02	1.0

'MID' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	1.0	.32	.43	.45	.41	.38	.26	.20	.18
VM	.32	1.0	.30	.38	.21	.37	.24	.15	.15
NF	.43	.30	1.0	.33	.34	.41	.43	.10	.21
SR	.45	.38	.33	1.0	.34	.34	.34	.07	.23
FG	.41	.21	.34	.34	1.0	.37	.19	.02	.15
WG	.38	.37	.41	.34	.37	1.0	.33	.10	.22
PS	.26	.24	.43	.34	.19	.33	1.0	.14	.25
DS	.20	.15	.10	.07	.02	.10	.14	1.0	.16
FS	.18	.15	.21	.23	.15	.22	.25	.16	1.0

'HIGH' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	1.0	.41	.35	.55	.46	.39	.39	.09	.12
VM	.41	1.0	.30	.26	.43	.34	.21	.18	.13
NF	.35	.30	1.0	.38	.37	.37	.38	.15	.27
SR	.55	.26	.38	1.0	.51	.27	.34	.07	.16
FG	.46	.43	.37	.51	1.0	.41	.23	.01	.16
WG	.39	.34	.37	.27	.41	1.0	.21	.17	.06
PS	.39	.21	.38	.34	.23	.21	1.0	.16	.07
DS	.09	.18	.15	.07	.10	.17	.16	1.0	.17
FS	.12	.13	.27	.16	.16	.06	.07	.17	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	1.0	.46	.50	.41	.43	.43	.29	.13	-.04
VM	.46	1.0	.31	.34	.20	.29	.21	.18	.04
NF	.50	.31	1.0	.34	.29	.36	.54	.32	.24
SR	.41	.34	.34	1.0	.31	.36	.32	.23	.21
FG	.43	.2	.29	.31	1.0	.42	.24	.13	.07
WG	.43	.29	.36	.36	.42	1.0	.33	.21	.15
PS	.29	.21	.54	.32	.24	.33	1.0	.29	.21
WF	.13	.18	.32	.23	.13	.21	.29	1.0	.21
FS	-.04	.04	.24	.21	.07	.15	.21	.21	1.0

'MID' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	1.0	.41	.43	.47	.44	.43	.33	.27	.26
VM	.41	1.0	.34	.36	.31	.45	.23	.37	.26
NF	.43	.34	1.0	.39	.41	.44	.51	.44	.27
SR	.47	.36	.39	1.0	.39	.33	.39	.22	.24
FG	.44	.31	.41	.39	1.0	.44	.29	.17	.15
WG	.43	.45	.44	.33	.44	1.0	.39	.32	.20
PS	.33	.23	.51	.39	.29	.39	1.0	.33	.22
WF	.27	.37	.44	.22	.17	.32	.33	1.0	.35
FS	.26	.26	.27	.24	.15	.20	.22	.35	1.0

'HIGH' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	1.0	.31	.42	.49	.37	.31	.36	.10	.16
VM	.31	1.0	.42	.33	.28	.42	.43	.29	.20
NF	.42	.42	1.0	.25	.26	.40	.42	.21	.28
SR	.49	.33	.25	1.0	.44	.36	.31	.20	.25
FG	.37	.28	.26	.44	1.0	.21	.19	.02	.26
WG	.31	.42	.40	.36	.21	1.0	.35	.19	.25
PS	.36	.43	.42	.31	.19	.35	1.0	.31	.33
WF	.10	.29	.21	.20	.02	.19	.31	1.0	.44
FS	.16	.20	.28	.25	.26	.25	.33	.44	1.0

APPENDIX 6.7 CONTINUED.
INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS.

'LOW' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	1.0	.39	.46	.39	.26	.39	.35	.09	.13
VM	.39	1.0	.31	.31	.07	.36	.32	.31	.12
NF	.46	.31	1.0	.24	.35	.41	.52	.13	.18
SR	.39	.31	.24	1.0	.36	.36	.36	.09	.08
FG	.26	.07	.35	.36	1.0	.30	.22	-.11	-.01
WG	.39	.36	.41	.36	.30	1.0	.43	.07	.06
PS	.35	.32	.52	.36	.22	.43	1.0	.22	.28
WF	.09	.31	.13	.09	-.11	.07	.22	1.0	-.08
DS	.13	.12	.18	.08	-.01	.06	.28	-.08	1.0

'MID' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	1.0	.36	.40	.45	.42	.41	.28	.18	.17
VM	.36	1.0	.30	.32	.29	.39	.19	.26	.10
NF	.40	.30	1.0	.32	.34	.37	.49	.36	.15
SR	.45	.32	.32	1.0	.32	.31	.33	.19	.08
FG	.42	.29	.34	.32	1.0	.41	.27	.11	.05
WG	.41	.39	.37	.31	.41	1.0	.36	.27	.12
PS	.28	.19	.49	.33	.27	.36	1.0	.32	.08
WF	.18	.26	.36	.19	.11	.27	.32	1.0	.12
DS	.17	.10	.15	.08	.05	.12	.08	.12	1.0

'HIGH' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	1.0	.47	.50	.50	.57	.37	.40	.28	.30
VM	.47	1.0	.41	.35	.42	.49	.31	.29	.27
NF	.50	.41	1.0	.41	.35	.52	.42	.43	.19
SR	.50	.35	.41	1.0	.54	.27	.36	.18	.04
FG	.57	.42	.35	.54	1.0	.37	.20	.27	.06
WG	.37	.49	.52	.27	.37	1.0	.21	.30	.15
PS	.40	.31	.42	.36	.20	.21	1.0	.22	.24
WF	.28	.29	.43	.18	.27	.30	.22	1.0	.05
DS	.30	.27	.19	.04	.06	.15	.24	.05	1.0

APPENDIX 6.8
TEST RELIABILITIES IN LOW, MID, AND HIGH ABILITY GROUPS SELECTED ACCORDING TO IQ ON EACH VARIABLE

SELECTOR: RAVEN'S MATRICES

	LOW	MID	HIGH
VM	.94	.95	.95
NF	.95	.96	.97
SR	.73	.82	.82
FG	.69	.74	.59
WG	.83	.74	.65
PS	.95	.95	.96
WF	.96	.96	.96
DS	.86	.85	.85
FS	.73	.83	.82
MEAN	.85	.87	.84

SELECTOR: VERBAL MEANING

	LOW	MID	HIGH
RPM	.94	.91	.90
NF	.97	.95	.96
SR	.83	.81	.82
FG	.82	.70	.69
WG	.80	.78	.54
PS	.96	.95	.96
WF	.95	.96	.97
DS	.85	.86	.87
FS	.84	.77	.86
MEAN	.88	.86	.84

SELECTOR: NUMBER FACILITY

	LOW	MID	HIGH
RPM	.93	.91	.89
VM	.94	.95	.93
SR	.83	.83	.71
FG	.82	.68	.64
WG	.88	.72	.64
PS	.97	.94	.95
WF	.96	.95	.96
DS	.88	.85	.85
FS	.74	.82	.77
MEAN	.88	.85	.82

SELECTOR: SPATIAL RELATIONS

	LOW	MID	HIGH
RPM	.94	.91	.85
VM	.95	.94	.94
NF	.96	.96	.95
FG	.82	.68	.68
WG	.84	.77	.60
PS	.96	.95	.95
WF	.96	.96	.96
DS	.89	.85	.85
FS	.83	.80	.82
MEAN	.91	.87	.84

SELECTOR: FIGURE GROUPING

	LOW	MID	HIGH
RPM	.94	.91	.89
VM	.93	.95	.94
NF	.95	.96	.96
SR	.85	.80	.85
WG	.84	.75	.70
PS	.95	.95	.96
WF	.96	.96	.95
DS	.85	.86	.84
FS	.80	.82	.82
MEAN	.90	.88	.88

SELECTOR: WORD GROUPING

	LOW	MID	HIGH
RPM	.92	.92	.85
VM	.95	.94	.95
NF	.98	.95	.95
SR	.83	.82	.8
FG	.76	.71	.68
PS	.95	.96	.95
WF	.95	.96	.97
DS	.84	.87	.82
FS	.77	.8	.86
MEAN	.88	.88	.87

APPENDIX 6.8 CONTINUED.
TEST RELIABILITIES IN LOW, MID, AND HIGH ABILITY GROUPS SELECTED ACCORDING TO IQ ON EACH VARIABLE

SELECTOR: PERCEPTUAL SPEED

	LOW	MID	HIGH
RPM	.94	.91	.90
VM	.94	.95	.93
NF	.95	.96	.96
SR	.86	.80	.77
FG	.79	.70	.73
WG	.77	.78	.67
WF	.94	.96	.96
DS	.88	.85	.85
FS	.82	.80	.82
MEAN	.88	.86	.84

SELECTOR: WORD FLUENCY

	LOW	MID	HIGH
RPM	.94	.91	.91
VM	.95	.94	.94
NF	.97	.95	.95
SR	.87	.81	.80
FG	.75	.73	.80
WG	.85	.77	.62
PS	.96	.95	.96
DS	.87	.85	.86
FS	.83	.79	.79
MEAN	.89	.86	.85

SELECTOR: DIGIT SPAN

	LOW	MID	HIGH
RPM	.93	.92	.90
VM	.92	.95	.95
NF	.95	.96	.96
SR	.81	.83	.83
FG	.72	.77	.61
WG	.76	.79	.77
PS	.96	.96	.95
WF	.96	.96	.96
FS	.82	.81	.82
MEAN	.87	.88	.86

SELECTOR: FIGURES OF SPEECH

	LOW	MID	HIGH
RPM	.93	.92	.90
VM	.95	.94	.94
NF	.97	.96	.94
SR	.86	.80	.82
FG	.77	.75	.68
WG	.85	.77	.55
PS	.95	.96	.95
WF	.97	.96	.96
DS	.86	.85	.87
MEAN	.90	.88	.85

APPENDIX 6.9

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS SELECTED ACCORDING TO IQ ON EACH OF THE TEN SUBTESTS, AFTER CORRECTING FOR UNRELIABILITY.

[THE TEST USED AS THE SELECTOR IS EXCLUDED FROM EACH MATRIX.]

'LOW' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	*	.35	.41	.22	.39	.23	.44	.12	.07
NF	.35	*	.25	.21	.17	.43	.44	.26	.31
SR	.41	.25	*	.29	.16	.10	.15	.04	.15
FG	.22	.21	.29	*	.42	.11	.14	.24	.15
WG	.39	.17	.16	.42	*	.23	.32	.14	.27
PS	.23	.43	.10	.11	.23	*	.43	.12	.21
WF	.44	.44	.15	.14	.32	.43	*	.13	.22
DS	.12	.26	.04	.24	.14	.12	.13	*	.07
FS	.07	.31	.15	.15	.27	.21	.22	.07	*

'MID' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	*	.25	.28	.17	.41	.20	.24	.09	.25
NF	.25	*	.30	.38	.44	.44	.33	.05	.26
SR	.28	.30	*	.30	.32	.39	.18	0.0	.29
FG	.17	.38	.30	*	.40	.27	.07	-.08	.09
WG	.41	.44	.32	.40	*	.35	.22	.01	.20
PS	.20	.44	.39	.27	.35	*	.24	.12	.24
WF	.24	.33	.18	.07	.22	.24	*	.07	.40
DS	.09	.05	0.0	-.08	.01	.12	.07	*	.13
FS	.25	.26	.29	.09	.20	.24	.40	.13	*

'HIGH' ABILITY GROUP SELECTED BY RAVEN'S MATRICES IQ

	VM	NF	SR	FG	WG	PS	WF	DS	FS
VM	*	.17	.21	.35	.27	.08	.32	.06	.14
NF	.17	*	.21	.09	.34	.52	.29	.14	.26
SR	.22	.21	*	.61	.36	.24	.20	.00	.16
FG	.35	.09	.61	*	.42	.03	.09	-.22	.21
WG	.27	.34	.36	.42	*	.44	.32	.10	.16
PS	.08	.52	.24	.03	.44	*	.33	.15	.20
WF	.32	.29	.20	.09	.32	.33	*	-.07	.38
DS	.06	.14	.00	-.22	.11	.15	-.07	*	.29
FS	.14	.26	.16	.21	.16	.20	.37	.29	*

APPENDIX 6.9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	*	.41	.53	.44	.54	.36	.22	.21	.15
NF	.41	*	.38	.44	.39	.61	.30	.22	.23
SR	.53	.38	*	.44	.54	.39	.15	.03	.16
FG	.44	.44	.44	*	.55	.44	.08	-.01	.17
WG	.54	.39	.54	.55	*	.40	.29	.02	.23
PS	.36	.61	.39	.44	.40	*	.19	.13	.19
WF	.22	.30	.15	.08	.29	.19	*	-.02	.47
DS	.21	.22	.03	-.01	.02	.13	-.02	*	.19
FS	.15	.23	.16	.17	.23	.19	.47	.19	*

'MID' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	*	.43	.44	.44	.38	.30	.09	.23	.13
NF	.43	*	.33	.4	.41	.49	.31	.16	.19
SR	.44	.33	*	.41	.30	.38	.17	.12	.25
FG	.44	.4	.41	*	.45	.23	.05	.05	.09
WG	.38	.41	.3	.45	*	.41	.22	.15	.16
PS	.30	.49	.38	.23	.41	*	.31	.18	.24
WF	.09	.31	.17	.05	.22	.31	*	.11	.29
DS	.23	.16	.12	.05	.15	.18	.11	*	.11
FS	.13	.19	.25	.09	.16	.24	.29	.11	*

'HIGH' ABILITY GROUP SELECTED BY VERBAL MEANING IQ

	RPM	NF	SR	FG	WG	PS	WF	DS	FS
RPM	*	.40	.44	.55	.32	.27	.22	.01	.23
NF	.40	*	.26	.25	.57	.39	.41	.17	.44
SR	.44	.26	*	.50	.21	.28	.14	-.08	.23
FG	.55	.25	.50	*	.35	.23	.21	.01	.27
WG	.32	.57	.21	.35	*	.23	.19	.12	.22
PS	.27	.39	.28	.23	.23	*	.28	.21	.26
WF	.22	.41	.14	.21	.19	.28	*	.03	.36
DS	.01	.17	-.08	.01	.12	.21	.03	*	.27
FS	.23	.44	.23	.27	.22	.26	.36	.27	*

APPENDIX 6.9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	*	.38	.18	.19	.21	.06	.05	.23	.19
VM	.38	*	.29	.25	.34	.23	.35	.20	.12
SR	.17	.29	*	.46	.23	.32	.11	.07	.07
FG	.19	.25	.46	*	.50	.13	.14	-.05	.21
WG	.21	.34	.23	.50	*	.30	.22	.00	.08
PS	.06	.23	.32	.13	.30	*	.36	.16	.16
WF	.05	.35	.11	.14	.22	.36	*	.22	.21
DS	.23	.20	.07	-.05	.00	.16	.22	*	.15
FS	.19	.12	.07	.21	.08	.16	.21	.15	*

'MID' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	*	.38	.53	.57	.48	.24	.13	.21	.12
VM	.38	*	.40	.31	.48	.22	.20	.14	.17
SR	.53	.40	*	.41	.37	.30	.14	.12	.21
FG	.57	.31	.41	*	.42	.22	.03	.07	.08
WG	.48	.48	.37	.42	*	.30	.19	.24	.21
PS	.24	.22	.30	.22	.30	*	.26	.15	.23
WF	.13	.20	.14	.03	.19	.26	*	.10	.32
DS	.21	.14	.12	.07	.24	.16	.10	*	.14
FS	.12	.17	.21	.08	.21	.23	.32	.14	*

'HIGH' ABILITY GROUP SELECTED BY NUMBER FACILITY IQ

	RPM	VM	SR	FG	WG	PS	WF	DS	FS
RPM	*	.31	.52	.29	.32	.35	.15	.07	.31
VM	.31	*	.31	.28	.43	.08	.48	.15	.48
SR	.52	.31	*	.50	.49	.45	.30	-.11	.49
FG	.29	.28	.50	*	.50	.22	.09	-.13	.2
WG	.32	.43	.49	.50	*	.44	.38	-.15	.33
PS	.35	.08	.45	.22	.44	*	.13	.09	.17
WF	.15	.48	.30	.09	.38	.13	*	-.09	.55
DS	.07	.15	-.11	-.13	-.15	.09	-.09	*	.17
FS	.31	.48	.49	.20	.33	.17	.55	.17	*

APPENDIX 6.9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	*	.38	.29	.35	.28	.09	.10	.30	.12
VM	.38	*	.34	.17	.57	.13	.31	.20	.12
NF	.29	.34	*	.38	.26	.46	.26	.26	.18
FG	.35	.17	.38	*	.31	.25	.10	.03	.14
WG	.28	.57	.26	.31	*	.30	.34	.20	.32
PS	.09	.13	.46	.25	.30	*	.26	.22	.13
WF	.10	.31	.26	.10	.34	.26	*	.17	.24
DS	.30	.20	.26	.03	.20	.22	.17	*	.27
FS	.12	.12	.18	.14	.32	.13	.24	.27	*

'MID' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	*	.38	.48	.40	.44	.33	.17	.15	.18
VM	.38	*	.34	.28	.41	.22	.31	.17	.24
NF	.48	.34	*	.31	.48	.51	.35	.19	.23
FG	.40	.28	.31	*	.48	.17	.02	.05	.03
WG	.44	.41	.48	.48	*	.38	.28	.12	.15
PS	.33	.22	.51	.17	.38	*	.33	.18	.25
WF	.17	.31	.35	.02	.28	.33	*	.11	.38
DS	.15	.17	.19	.05	.12	.18	.11	*	.13
FS	.18	.24	.23	.03	.15	.25	.38	.13	*

'HIGH' ABILITY GROUP SELECTED BY SPATIAL RELATIONS IQ

	RPM	VM	NF	FG	WG	PS	WF	DS	FS
RPM	*	.28	.29	.66	.57	.35	.18	.25	.05
VM	.28	*	.19	.35	.33	.27	.21	.10	.12
NF	.29	.19	*	.50	.52	.40	.42	.08	.43
FG	.66	.35	.50	*	.59	.33	.28	.06	.35
WG	.57	.33	.52	.59	*	.44	.13	.15	.21
PS	.35	.27	.40	.33	.44	*	.19	.15	.22
WF	.18	.21	.42	.28	.13	.19	*	-.05	.37
DS	.25	.10	.08	.06	.16	.15	-.05	*	.20
FS	.05	.12	.43	.35	.21	.22	.37	.20	1.22

APPENDIX 6.9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	*	.25	.35	.32	.24	.23	.27	.22	.06
VM	.25	*	.37	.30	.46	.31	.32	.18	.09
NF	.35	.37	*	.45	.44	.47	.53	-.05	.28
SR	.32	.30	.45	*	.22	.40	.28	.12	.30
WG	.24	.46	.44	.22	*	.32	.38	.17	.20
PS	.23	.31	.47	.40	.32	*	.55	.04	.28
WF	.27	.32	.53	.28	.38	.55	*	.18	.26
DS	.22	.18	-.05	.12	.17	.04	.18	*	.01
FS	.06	.09	.28	.30	.20	.28	.26	.01	*

'MID' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	*	.41	.40	.46	.43	.35	.17	.22	.19
VM	.41	*	.26	.35	.43	.22	.31	.15	.24
NF	.40	.26	*	.24	.38	.49	.35	.27	.27
SR	.46	.35	.24	*	.35	.36	.17	.05	.23
WG	.43	.43	.38	.35	*	.39	.29	.15	.25
PS	.35	.22	.49	.36	.39	*	.25	.24	.25
WF	.17	.31	.35	.17	.29	.25	*	.07	.40
DS	.22	.15	.27	.05	.15	.24	.07	*	.21
FS	.19	.24	.27	.23	.25	.25	.40	.21	*

'HIGH' ABILITY GROUP SELECTED BY FIGURE GROUPING IQ

	RPM	VM	NF	SR	WG	PS	WF	DS	FS
RPM	*	.33	.43	.58	.46	.17	.25	.34	.39
VM	.33	*	.40	.34	.42	.23	.38	.37	.33
NF	.43	.40	*	.38	.44	.49	.33	.22	.37
SR	.58	.34	.38	*	.47	.38	.36	.33	.36
WG	.46	.42	.44	.47	*	.33	.21	.15	.21
PS	.17	.23	.49	.38	.33	*	.33	.29	.33
WF	.25	.38	.33	.36	.21	.33	*	.19	.45
DS	.34	.37	.22	.33	.15	.29	.19	*	.23
FS	.39	.33	.37	.36	.21	.33	.45	.23	*

APPENDIX 6·9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	*	.45	.33	.46	.44	.21	.16	.25	.17
VM	.45	*	.23	.54	.22	.16	.38	.06	.24
NF	.33	.23	*	.30	.55	.45	.22	.29	.23
SR	.46	.54	.30	*	.36	.30	.29	.06	.38
FG	.44	.22	.55	.36	*	.40	.15	.23	.22
PS	.21	.16	.45	.30	.40	*	.30	.18	.23
WF	.16	.38	.22	.29	.15	.3	*	.15	.27
DS	.25	.06	.29	.06	.23	.18	.15	*	.09
FS	.17	.24	.23	.38	.22	.23	.27	.09	*

'MID' ABILITY GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	*	.36	.46	.42	.42	.27	.16	.21	.18
VM	.36	*	.34	.32	.28	.26	.25	.17	.15
NF	.46	.34	*	.35	.29	.49	.38	.18	.27
SR	.42	.32	.35	*	.42	.40	.18	.11	.24
FG	.42	.28	.29	.42	*	.17	.05	.07	.06
PS	.27	.26	.49	.40	.17	*	.34	.16	.29
WF	.16	.25	.38	.18	.05	.34	*	.06	.36
DS	.21	.17	.18	.11	.07	.16	.06	*	.24
FS	.18	.15	.27	.24	.06	.29	.36	.24	*

'HIGH' ABILITY GROUP SELECTED BY WORD GROUPING IQ

	RPM	VM	NF	SR	FG	PS	WF	DS	FS
RPM	*	.20	.31	.66	.49	.42	.12	.13	.14
VM	.20	*	.18	.14	.13	.03	.29	.22	.33
NF	.31	.18	*	.29	.33	.47	.34	.08	.33
SR	.66	.14	.29	*	.53	.26	.13	.00	.20
FG	.49	.13	.33	.53	*	.26	.13	-.33	.4
PS	.42	.03	.47	.26	.26	*	.08	.23	.05
WF	.12	.29	.34	.13	.13	.08	*	.13	.40
DS	.13	.22	.08	.00	-.33	.23	.13	*	.09
FS	.14	.33	.33	.20	.41	.05	.40	.09	*

APPENDIX 6·9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	*	.47	.31	.15	.32	.27	.24	.23	.21
VM	.47	*	.47	.32	.35	.41	.29	.15	.29
NF	.31	.47	*	.24	.35	.29	.43	.23	.31
SR	.15	.32	.24	*	.33	.11	.22	.08	.21
FG	.32	.35	.35	.33	*	.48	.31	.06	.17
WG	.27	.41	.29	.11	.48	*	.31	.03	.42
WF	.24	.29	.43	.22	.31	.31	*	.08	.34
DS	.23	.15	.23	.08	.06	.03	.08	*	.33
FS	.21	.29	.31	.21	.17	.42	.34	.33	*

'MID' ABILITY GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	*	.39	.42	.54	.57	.45	.15	.18	.16
VM	.39	*	.30	.34	.29	.46	.28	.12	.16
NF	.42	.30	*	.31	.35	.41	.30	.11	.23
SR	.54	.34	.31	*	.43	.39	.20	.05	.26
FG	.57	.29	.35	.43	*	.48	.07	.02	.14
WG	.45	.46	.41	.39	.48	*	.28	.11	.18
WF	.15	.28	.30	.20	.07	.28	*	.07	.35
DS	.18	.12	.11	.05	.02	.11	.07	*	.08
FS	.16	.16	.23	.26	.14	.18	.35	.08	*

'HIGH' ABILITY GROUP SELECTED BY PERCEPTUAL SPEED IQ

	RPM	VM	NF	SR	FG	WG	WF	DS	FS
RPM	*	.32	.34	.56	.34	.47	.06	.12	.16
VM	.32	*	.12	.41	.34	.42	.36	.22	.29
NF	.34	.12	*	.22	.38	.44	.14	.09	.19
SR	.56	.41	.22	*	.53	.41	-.02	.00	.22
FG	.34	.34	.38	.53	*	.4	-.03	.00	.19
WG	.47	.42	.44	.41	.4	*	-.03	.18	.05
WF	.06	.36	.14	-.02	-.03	-.03	*	-.03	.34
DS	.12	.22	.09	.00	.00	.18	-.03	*	.19
FS	.16	.29	.19	.22	.19	.05	.34	.19	*

APPENDIX 6·9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	*	.52	.49	.40	.51	.46	.39	.26	.09
VM	.52	*	.29	.29	.40	.47	.18	.11	.19
NF	.49	.29	*	.29	.46	.32	.65	.40	.12
SR	.40	.29	.29	*	.48	.33	.39	.14	.17
FG	.51	.40	.46	.48	*	.54	.55	.14	.08
WG	.46	.47	.32	.33	.54	*	.51	.12	.01
PS	.39	.18	.65	.39	.55	.51	*	.24	.07
DS	.26	.11	.40	.14	.14	.12	.24	*	.02
FS	.09	.19	.12	.17	.08	.01	.07	.02	*

'MID' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	*	.35	.46	.52	.50	.46	.28	.23	.21
VM	.35	*	.31	.43	.25	.44	.26	.17	.17
NF	.46	.31	*	.38	.41	.48	.46	.11	.24
SR	.52	.43	.38	*	.44	.43	.39	.09	.29
FG	.50	.25	.41	.44	*	.49	.23	.02	.20
WG	.46	.44	.48	.43	.49	*	.39	.13	.28
PS	.28	.26	.46	.39	.23	.39	*	.15	.29
DS	.23	.17	.11	.09	.02	.13	.15	*	.19
FS	.21	.17	.24	.29	.20	.28	.29	.19	*

'HIGH' ABILITY GROUP SELECTED BY WORD FLUENCY IQ

	RPM	VM	NF	SR	FG	WG	PS	DS	FS
RPM	*	.44	.37	.64	.54	.52	.42	.10	.14
VM	.44	*	.31	.29	.50	.44	.22	.21	.15
NF	.37	.31	*	.43	.43	.48	.39	.17	.31
SR	.64	.30	.43	*	.64	.38	.39	.08	.21
FG	.54	.50	.43	.64	*	.58	.27	.12	.20
WG	.52	.44	.48	.38	.58	*	.27	.24	.09
PS	.42	.22	.39	.39	.27	.27	*	.18	.08
DS	.10	.21	.17	.08	.12	.24	.18	*	.20
FS	.14	.15	.31	.21	.20	.09	.08	.20	*

APPENDIX 6·9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	*	.49	.53	.47	.53	.51	.31	.13	-.04
VM	.49	*	.33	.40	.25	.35	.22	.19	.05
NF	.53	.33	*	.39	.35	.42	.56	.33	.27
SR	.47	.40	.39	*	.41	.46	.37	.26	.25
FG	.53	.25	.35	.41	*	.57	.29	.15	.09
WG	.51	.35	.42	.46	.57	*	.39	.24	.19
PS	.31	.22	.56	.37	.29	.39	*	.30	.23
WF	.13	.19	.33	.26	.15	.24	.30	*	.23
FS	-.04	.05	.27	.25	.09	.19	.23	.23	*

'MID' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	*	.43	.45	.53	.52	.50	.35	.28	.30
VM	.43	*	.35	.41	.37	.52	.25	.39	.30
NF	.45	.35	*	.43	.48	.51	.53	.45	.31
SR	.53	.41	.43	*	.49	.41	.44	.25	.30
FG	.52	.37	.48	.49	*	.56	.34	.20	.19
WG	.50	.52	.51	.41	.56	*	.44	.36	.25
PS	.35	.25	.53	.44	.34	.44	*	.35	.25
WF	.28	.39	.45	.25	.20	.36	.35	*	.40
FS	.30	.30	.31	.30	.19	.25	.25	.40	*

'HIGH' ABILITY GROUP SELECTED BY DIGIT SPAN IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	FS
RPM	*	.34	.45	.57	.50	.38	.38	.11	.19
VM	.34	*	.44	.37	.37	.49	.45	.31	.22
NF	.45	.44	*	.28	.33	.46	.44	.22	.32
SR	.57	.37	.28	*	.61	.45	.35	.22	.30
FG	.50	.37	.33	.61	*	.31	.25	.03	.37
WG	.38	.49	.46	.45	.31	*	.40	.22	.31
PS	.38	.45	.44	.35	.25	.40	*	.32	.37
WF	.11	.31	.22	.22	.03	.22	.32	*	.50
FS	.19	.22	.32	.30	.37	.31	.37	.50	*

APPENDIX 6·9 CONTINUED.

INTER-TEST CORRELATIONS IN 'LOW', 'MID' AND 'HIGH' ABILITY GROUPS, CORRECTED FOR UNRELIABILITY.

'LOW' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	*	.41	.49	.43	.31	.43	.38	.10	.14
VM	.41	*	.32	.34	.08	.40	.34	.32	.13
NF	.49	.32	*	.27	.40	.45	.54	.14	.20
SR	.43	.34	.27	*	.44	.42	.40	.10	.09
FG	.31	.08	.41	.44	*	.37	.25	-.12	-.01
WG	.43	.40	.45	.42	.37	*	.48	.07	.07
PS	.38	.34	.54	.40	.25	.48	*	.23	.31
WF	.10	.32	.14	.10	-.12	.07	.23	*	-.09
DS	.14	.13	.20	.09	-.01	.07	.31	-.09	*

'MID' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	*	.39	.43	.53	.51	.48	.30	.19	.19
VM	.39	*	.32	.37	.35	.45	.20	.28	.12
NF	.43	.32	*	.37	.40	.42	.51	.37	.16
SR	.53	.37	.37	*	.41	.40	.37	.22	.09
FG	.51	.35	.40	.41	*	.54	.32	.13	.07
WG	.48	.45	.42	.40	.54	*	.42	.32	.15
PS	.30	.20	.51	.37	.32	.42	*	.34	.09
WF	.19	.28	.37	.22	.13	.32	.34	*	.13
DS	.19	.12	.16	.09	.07	.15	.09	.13	*

'HIGH' ABILITY GROUP SELECTED BY FIGURES OF SPEECH IQ

	RPM	VM	NF	SR	FG	WG	PS	WF	DS
RPM	*	.51	.54	.58	.73	.53	.43	.30	.34
VM	.51	*	.44	.40	.52	.67	.33	.30	.30
NF	.54	.44	*	.47	.44	.72	.45	.45	.21
SR	.58	.40	.47	*	.72	.40	.41	.20	.04
FG	.73	.52	.44	.72	*	.60	.25	.34	.07
WG	.53	.68	.72	.4	.60	*	.29	.41	.22
PS	.43	.33	.45	.41	.25	.29	*	.23	.26
WF	.30	.30	.45	.20	.34	.41	.23	*	.05
DS	.34	.30	.21	.04	.07	.22	.26	.05	*

APPENDIX 11

(Appendix for chapter 11).

11.1 Correlations of IQ variables with parameters of the letter discrimination task in the full sample.

APPENDIX 11.1

All Subjects [N=115].
Correlations of IQ Variables with Letter Discrimination Time in 'Physical Identity' and 'Name Identity' Conditions, with RT across conditions, and with Speed of Lexical Access as measured by the subtraction of P from N.

	Physical ID	Name ID	Average RT	N - P
Raven's Matrices	-.21	-.29	-.28	-.17
Spatial Relations	-.13	-.18	-.17	-.16
Figure Grouping	-.19	-.20	-.23	-.10
Perceptual Speed	-.28	-.37	-.39	-.18
Verbal Meaning	-.04	-.10	-.10	-.13
Word Grouping	-.05	-.17	-.16	-.18
Word Fluency	-.08	-.24	-.18	-.25
Figures of Speech	-.03	-.24	-.16	-.27
Number Facility	-.28	-.32	-.35	-.10
Digit Span	.06	-.01	.04	-.05
Average	-.12	-.21	-.20	-.16